

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 04:44:47 : Search time 2160.68 Seconds
(without alignments)
3631.936 Million cell updates/sec

Title: US-09-761-466-1
Perfect score: 375
Sequence: 1 aggcctccgcacccctcctc.....aggtcttcacacgcacac 375

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 30

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: qb_ba.*
- 2: qb_hrq.*
- 3: qb_in.*
- 4: qb_om.*
- 5: qb_ov.*
- 6: qb_pat.*
- 7: qb_ph.*
- 8: qb_pl.*
- 9: qb_pr.*
- 10: qb_ro.*
- 11: qb_sts.*
- 12: qb_sy.*
- 13: qb_un.*
- 14: qb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_mu.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_to.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htq_hum.*
- 31: em_htq_inv.*
- 32: em_htq_other.*
- 33: em_htqo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

295	78.7	136657	2	AL669924	Homo sapi
295	78.7	137545	2	AC008412	Homo sapi
295	78.7	144762	2	AL669821	Homo sapi
31	8.3	79364	2	AC102990	Rattus no

ALIGNMENTS

AL669924	136657	bp	1NA	linear	HIG 18-JAN-2002
Homo sapiens chromosome 6 clone: XXbac-25c14	***	7	unordered pieces	***	SEQUENCING IN
PROGRESS	***	7	unordered pieces		
AL669924	GI:18375903				
HTG: HTGS_PHASE1: HTGS_ACTIVEFIN: HTGS_DRAFT: HTGS_FULLTOP.					
Human					
Homo sapiens					
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:					
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo					
1 (sites)					
Peck, A.					
Direct Submission					
Submitted (17-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,					
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:					
humquery@sanger.ac.uk					
On Jan 25, 2002 this sequence version replaced gi:18250865.					
***** Genome Center					
Center: Wellcome Trust Sanger Institute					
Center code: SC					
Web site: http://www.sanger.ac.uk					
Contact: humquery@sanger.ac.uk					
***** Project Information					
Center project name: bp9256J14					
***** Summary Statistics					
Assembly program: XGAP4; version 4.5					
Chemistry: dye-terminator Big Dye; 100% of reads					
Consensus quality: 134624 bases at least 40					
Consensus quality: 135417 bases at least 20					
Consensus quality: 135772 bases at least 10					
Insert size: 136057; sum-of: contigs					
Insert size: 115180; 6.1% error: aaadose-1f					
Quality coverage: 7.81x in 920 bases; sum-of: contigs Quality					
coverage: 9.24x in 920 bases; aaadose-1f					

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	7523;	contig of 7523 bp in length
7524	7623;	gap of 100 bp
7524	41360;	contig of 43747 bp in length
41361	41460;	gap of 100 bp
41461	47515;	contig of 6265 bp in length
47516	47615;	gap of 100 bp
47616	52829;	contig of 5214 bp in length
52830	52929;	gap of 100 bp
52930	90494;	contig of 37664 bp in length
90494	90593;	gap of 100 bp
90594	109805;	contig of 19212 bp in length
109806	109905;	gap of 100 bp
109906	136657;	contig of 2772 bp in length.

Location/Qualifiers
1. 136657
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXbac-25c14"

FEATURES
source

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		On Jul 18, 2000 this sequence version replaced gi:7708838.	
-----Genome Center			
Center: Joint Genome Institute			
Center Code: JGI			
Web site: http://www.jgi.doe.gov			

Project Information			
Center Project Name: 293265			
Center clone name: CIT-HSPC_281H14			

Summary Statistics			
Consensus quality: 123959 bases at least Q40			
Consensus quality: 133998 bases at least Q30			
Consensus quality: 134736 bases at least Q20			
Estimated insert size: 136540; agarose-fp estimation			
Estimated insert size: 136595; sum-of-contigs estimation			
Quality coverage: 4.95 in Q20 bases; agarose-fp estimation			
Quality coverage: 4.91 in Q20 bases; sum-of-contigs estimation.			
* NOTE: This is a 'working draft' sequence. It currently			
* consists of 21 contigs. Gaps between the contigs			
* are represented as runs of N. The order of the pieces			
* is believed to be correct as given, however the sizes			
* of the gaps between them are based on estimates that have			
* provided by the submittor.			
* This sequence will be replaced			
* by the finished sequence as soon as it is available and			
* the accession number will be preserved.			
1 2162: contig of 2162 bp in length			
2163 2262: gap of unknown length			
2263 5223: contig of 2961 bp in length			
5224 5323: gap of unknown length			
5324 15199: contig of 9876 bp in length			
15200 15299: gap of unknown length			
15300 22312: contig of 7013 bp in length			
22313 22412: gap of unknown length			
22413 37107: contig of 14695 bp in length			
37108 43863: contig of 6656 bp in length			
43864 43963: gap of unknown length			
43964 46965: contig of 3002 bp in length			
46966 47065: gap of unknown length			
47066 49477: contig of 2412 bp in length			
49478 49577: gap of unknown length			
49578 54523: contig of 4946 bp in length			
54524 54623: gap of unknown length			
54624 68414: contig of 13791 bp in length			
68415 68514: gap of unknown length			
68515 72850: contig of 4336 bp in length			
72851 72950: gap of unknown length			
72951 88213: contig of 15263 bp in length			
88214 88313: gap of unknown length			
88314 98538: contig of 10225 bp in length			
98539 98638: gap of unknown length			
98639 107774: contig of 9136 bp in length			
107775 107874: gap of unknown length			
107875 115145: contig of 7271 bp in length			
115146 115245: gap of unknown length			
115246 118552: contig of 3307 bp in length			
118553 118652: gap of unknown length			
118653 128198: contig of 9546 bp in length			
128199 128298: gap of unknown length			
128299 130123: contig of 1825 bp in length			
130124 130223: gap of unknown length			
130224 134038: contig of 3815 bp in length			
134039 134138: gap of unknown length			
134139 136470: contig of 2332 bp in length			
136471 136570: gap of unknown length			
136571 137545: contig of 975 bp in length.			
Location/Qualifiers			
1. .137545			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			

COMMENT		Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
On Jul 18, 2000 this sequence version replaced gi:7708838.			
-----Genome Center			
Center: Joint Genome Institute			
Center Code: JGI			
Web site: http://www.jgi.doe.gov			

Project Information			
Center Project Name: 293265			
Center clone name: CIT-HSPC_281H14			

Summary Statistics			
Consensus quality: 123959 bases at least Q40			
Consensus quality: 133998 bases at least Q30			
Consensus quality: 134736 bases at least Q20			
Estimated insert size: 136540; agarose-fp estimation			
Estimated insert size: 136595; sum-of-contigs estimation			
Quality coverage: 4.95 in Q20 bases; agarose-fp estimation			
Quality coverage: 4.91 in Q20 bases; sum-of-contigs estimation.			
* NOTE: This is a 'working draft' sequence. It currently			
* consists of 21 contigs. Gaps between the contigs			
* are represented as runs of N. The order of the pieces			
* is believed to be correct as given, however the sizes			
* of the gaps between them are based on estimates that have			
* provided by the submittor.			
* This sequence will be replaced			
* by the finished sequence as soon as it is available and			
* the accession number will be preserved.			
1 2162: contig of 2162 bp in length			
2163 2262: gap of unknown length			
2263 5223: contig of 2961 bp in length			
5224 5323: gap of unknown length			
5324 15199: contig of 9876 bp in length			
15200 15299: gap of unknown length			
15300 22312: contig of 7013 bp in length			
22313 22412: gap of unknown length			
22413 37107: contig of 14695 bp in length			
37108 43863: contig of 6656 bp in length			
43864 43963: gap of unknown length			
43964 46965: contig of 3002 bp in length			
46966 47065: gap of unknown length			
47066 49477: contig of 2412 bp in length			
49478 49577: gap of unknown length			
49578 54523: contig of 4946 bp in length			
54524 54623: gap of unknown length			
54624 68414: contig of 13791 bp in length			
68415 68514: gap of unknown length			
68515 72850: contig of 4336 bp in length			
72851 72950: gap of unknown length			
72951 88213: contig of 15263 bp in length			
88214 88313: gap of unknown length			
88314 98538: contig of 10225 bp in length			
98539 98638: gap of unknown length			
98639 107774: contig of 9136 bp in length			
107775 107874: gap of unknown length			
107875 115145: contig of 7271 bp in length			
115146 115245: gap of unknown length			
115246 118552: contig of 3307 bp in length			
118553 118652: gap of unknown length			
118653 128198: contig of 9546 bp in length			
128199 128298: gap of unknown length			
128299 130123: contig of 1825 bp in length			
130124 130223: gap of unknown length			
130224 134038: contig of 3815 bp in length			
134039 134138: gap of unknown length			
134139 136470: contig of 2332 bp in length			
136471 136570: gap of unknown length			
136571 137545: contig of 975 bp in length.			
Location/Qualifiers			
1. .137545			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			

FEATURES		source	
1. .137545			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			

misc_feature		/clone_lib="CHORI-501"	
1. .7523		/note="assembly_fragment:01480	
fragment_chain:1		clone_end:T7	
vector_side:left		7624. .41360	
/note="assembly_fragment:01718		fragment_chain:1"	
41461. .47515		/note="assembly_fragment:00204	
/note="assembly_fragment:00204		fragment_chain:1"	
47616. .52829		/note="assembly_fragment:02002	
/note="assembly_fragment:02002		fragment_chain:1"	
52930. .90493		/note="assembly_fragment:00959	
/note="assembly_fragment:00959		fragment_chain:1"	
90594. .109805		/note="assembly_fragment:01522	
/note="assembly_fragment:01522		fragment_chain:1"	
109906. .136657		/note="assembly_fragment:01742	
/note="assembly_fragment:01742		fragment_chain:1"	
BASE COUNT 33294 a 32671 c 32974 g 37117 t 601 others			
ORIGIN			
Query Match 78.7%; Score 295; DB 2: Length 136657;			
Best Local Similarity 100.0%; Pred. NO. 1.3e-160;			
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY 1 agggcccccgcacccatccttgctcccccgcctctctccacccctcccgagcccccctaa 60			
Db 54587 AGGCCCGCGGCGCCATCGTGGCTCCGCCCTCTCTCCACCTCCGCGACCCCTAA 54528			
QY 61 agggcgcgcgggcccaagcgagggcgctgcgcctgaccccgagcggaagggccccagt 120			
Db 54527 AGGGCGGCGGCGGCGCAAGCGGCGCTGCGCTGACCGCGAGCGGAGGCGGCCAGT 54468			
QY 121 ctaggtcctaagcgggtggcgctctctttgacagcgcgctttggggacacacggggg 180			
Db 54467 CTAGGTCTCTAATGCGGCGGCGCTCTCTTTGACAGCGCGCGTTCGGGACACACGCGGG 54408			
QY 181 acgagagataaggtgacataccagacagatttggtgcgcgctgatactctctcccg 240			
Db 54407 ACGAGAGATAAGGTGACATACCAGAGCAGATTGGTGGCGCGCTGATCTCTCTCCG 54348			
QY 241 acaggaacgcggagctatttaaagaccctatcgattactttctctctcgga 295			
Db 54347 ACAGAAACGGGAGCTATTAAAGACCCATCGATTACATTATCTTTCTCTGGA 54293			
RESULT 2			
AC008412		137545 bp DNA linear HTG 18-JUL-2000	
LOCUS			
DEFINITION		Homo sapiens chromosome 5 clone CTC-281H14, WORKING DRAFT SEQUENCE,	
21 ordered pieces.			
AC008412			
VERSION		AC008412.5 GI:9255970	
KEYWORDS		HTG; HTGS_PHASE2; HTGS_DRAFT.	
SOURCE		human.	
ORGANISM		Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
1 (bases 1 to 137545)			
DOE Joint Genome Institute.			
TITLE		Sequencing of Human Chromosome 5	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 137545)	
AUTHORS		DOE Joint Genome Institute.	
DIRECT SUBMISSION			
TITLE		Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint	
JOURNAL			

chromosome "5"
/clone="CIC-281H14"
/clone_lib="Caltech human BAC library C"
BASE COUNT 37277 4 32408 C 31466 G 34494 T 2000 Others
ORIGIN

Query Match 78.7% Score 295; DB 2: Length 137545;
Best Local Similarity 100.0%; Pred. No. 1.3e+160;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agggcccccacacccatctactctcccccctctctccaccctccgagaccctaa 60
|||||
Db 28442 AGGACCCGCCACACCGTCATCGTCGCGCGCTTCCTCCACCCCTCCCGACCCCTAA 28501
QY 61 aagagacag 120
|||||
Db 28502 AGGAGCG 28561
QY 121 ctatctctaatcag 180
|||||
Db 28562 CTAATCTCTAATCGCGTGGGCTCTCTTTTGACAGCGCGCGCTTTGGGACACACGCGG 28621
QY 181 acag 240
|||||
Db 28622 ACAGAGAGATAGAGTACATAACAGAGAGAGATTTGGTGGCGCGCTGATACCTCTCTCT 28681
QY 241 acag 295
|||||
Db 28682 ACAAGAAACCGGACATATTAAGACAGAGATGATGATTTCTCTCTCTGGA 28746

RESULT 3
AL669821
LOCUS
DEFINITION Homo sapiens chromosome 6 clone XXbac-116A1, linear, HTG 31-JAN-2002
PROGRESS ***, 4 unordered pieces.

ACCESSION AL669821.4 31-18157151
VERSION
KEYWORDS HTG; HUGS_PHASE1; HUGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Mashreshi-Mohammadi,M.
Direct Submission
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
hampury@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jan 15, 2002 this sequence version replaced gi:18152648.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Contig code: SC
Web site: <http://www.sanger.ac.uk>
Contact: hampury@sanger.ac.uk
----- Project Information
Center project name: bpi16A1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 144176 bases at least Q40
Consensus quality: 144232 bases at least Q30
Consensus quality: 144259 bases at least Q20
Insert size: 144402; sum-of-contigs
Insert size: 167909; 14.3% error; adaroose-ip
Quality coverage: 10.72x in Q20 bases; sum-of-contigs Quality
coverage: 9.55x in Q20 bases; adaroose-ip

* NOTE: This is a 'working draft' sequence. It currently
consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 75928: contig of 75928 bp in length
* 75929 76028: gap of 100 bp
* 76029 116346: contig of 40346 bp in length
* 116347 116446: gap of 100 bp
* 116447 131258: contig of 14612 bp in length
* 131259 131358: gap of 100 bp
* 131359 144702: contig of 13444 bp in length.

FEATURES
source
1..144702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXbac-116A1"
/clone_lib="CHORI-11"

misc_feature
1..75928
/note="assembly fragment 1" 2
fragment_1(1..75928)
clone_end(1)

misc_feature
76029..116346
/note="assembly fragment 2" 3
fragment_2(1..40346)
fragment_end(1)

misc_feature
116447..131258
/note="assembly fragment 3" 4
fragment_3(1..14612)
fragment_end(1)

misc_feature
131359..144702
/note="assembly fragment 4" 5
fragment_4(1..13444)
fragment_end(1)
clone_end(1)
vector_start(1)
vector_end(1)
BASE COUNT 41517 A 33278 C 42869 G 39494 T 3000 Others
ORIGIN

Query Match 78.7% Score 295; DB 2: Length 144702;
Best Local Similarity 100.0%; Pred. No. 1.3e+160;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aagcccccacacccctctactctcccccctctctccaccctccgagaccctaa 60
|||||
Db 3010 AGGACCCGCCACACCGTCATCGTCGCGCGCTTCCTCCACCCCTCCCGACCCCTAA 3069

QY 61 aagagacag 120
|||||
Db 4070 AGGAGCG 4129

QY 121 ctatctctaatcag 180
|||||
Db 4140 CTATCTCTAATCGCGTGGGCTCTCTTTTGACAGCGCGCGCTTTGGGACACACGCGG 4199

QY 181 acag 240
|||||
Db 3190 ACAGAGAGATAGAGTACATAACAGAGAGAGATTTGGTGGCGCGCTGATACCTCTCTCT 3249

QY 241 acag 295
|||||
Db 3250 ACAGAGAGATAGAGTACATAACAGAGAGAGATTTGGTGGCGCGCTGATACCTCTCTCT 3304

RESULT 4
AC102990
LOCUS
DEFINITION AC102990.2 7604 bp, linear, HTG 20-JUN-92
PROGRESS ***, 51 unordered pieces.
ACCESSION AC102990
VERSION AC102990.2 01-17-92
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway 1st.

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 79364)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei-G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,S., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogoh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Thomas,N.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleciyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL

2 (bases 1 to 79364)

REFERENCE
AUTHORS

Worley,K.C.
Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062509.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHEJ
Center clone name: CH230-53024
----- Summary Statistics
findPhrapList
Assembly program: Phrap; version 0.990329First call to
Consensus quality: 57802 bases at least Q40
Consensus quality: 64028 bases at least Q30
Consensus quality: 67953 bases at least Q20
Estimated insert size: 35926; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
2616: contig of 2616 bp in length
2617: gap of unknown length
2717: contig of 3060 bp in length
5777: gap of unknown length
5777: contig of 1817 bp in length
7693: gap of unknown length
7694: contig of 1191 bp in length
7794: gap of unknown length
8985: contig of 1963 bp in length
9085: gap of unknown length
11047: contig of 1302 bp in length
11147: gap of unknown length
12449: contig of 1070 bp in length
12549: gap of unknown length
13619: contig of 1070 bp in length
13620: gap of unknown length
13719: contig of 2198 bp in length
15917: gap of unknown length
16017: contig of 2042 bp in length
18059: gap of unknown length
18159: contig of 1471 bp in length
19630: gap of unknown length
19631: contig of 1550 bp in length
19731: gap of unknown length
21280: contig of 1341 bp in length
21380: gap of unknown length
22721: contig of 1341 bp in length
22821: gap of unknown length
24148: contig of 1327 bp in length
24248: gap of unknown length
26033: contig of 1785 bp in length
26133: gap of unknown length
27521: contig of 1388 bp in length
27621: gap of unknown length
29570: contig of 1949 bp in length
29670: gap of unknown length
31004: contig of 1334 bp in length
31104: gap of unknown length
32567: contig of 1463 bp in length
32667: gap of unknown length
34343: contig of 1676 bp in length
34443: gap of unknown length
35645: contig of 1203 bp in length
35746: gap of unknown length
36828: contig of 1082 bp in length
36928: gap of unknown length
38702: contig of 1774 bp in length
38802: gap of unknown length
40269: contig of 1467 bp in length
40369: gap of unknown length
41638: contig of 1269 bp in length
41738: gap of unknown length
43021: contig of 1283 bp in length
43121: gap of unknown length
44403: contig of 1282 bp in length
44503: gap of unknown length
45796: contig of 1293 bp in length
45896: gap of unknown length
47159: contig of 1263 bp in length
47259: gap of unknown length
47260: contig of 1405 bp in length
48664: gap of unknown length
48764: contig of 1507 bp in length
50271: gap of unknown length
50371: contig of 1386 bp in length
51575: gap of unknown length
51758: contig of 1653 bp in length
53511: gap of unknown length
53610: contig of 1611 bp in length
55221: contig of 1611 bp in length


```
QY 361 ctgcacagcgacgc 375
|||||
Db 361 ctgcacagcgacgc 375

RESULT 2
AAS09960
ID AAS09960 standard; DNA; 1072 BP.
XX
AC AAS09960;
XX
DT 24-OCT-2001 (first entry)
XX
DE DNA encoding mCsx/Nkx2.5 homology domains A1 + A2, and intervening DNA.
XX
KW mCsx/Nkx2.5 homology domain; cardiac enhancer; cardiac cell;
KW cardiomyocyte induction; therapeutic; heart tissue; gene therapy;
KW mouse; ds.
XX
OS Mus musculus.
XX
PN WO200151006-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01511.
XX
PR 14-JAN-2000; 2000US-0176419.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lee IW, Izumo S;
XX
DR WPI; 2001-451809/48.
XX
CC The sequence represents the coding sequence of cardiac enhancer
CC mCsx/Nkx2.5 homology domains A1 + A2 and intervening sequence. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
SQ Sequence 1072 BP; 273 A; 290 C; 297 G; 212 T; 0 other;

Query Match 100.0%; Score 375; DB 22; Length 1072;
Best Local Similarity 100.0%; Pred. No. 9.3e-184;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agccccccgacccctacgtgctccgcgcgccttctctccacccctcccgagccctaa 60
|||||
Db 1 agccccccgacccctacgtgctccgcgcgccttctctccacccctcccgagccctaa 60

QY 61 agggcgcgggggcccaagcgagggcgctgcgtaccgccgagcggaagggccccagt 120
|||||
Db 61 agggcgcgggggcccaagcgagggcgctgcgtaccgccgagcggaagggccccagt 120

QY 121 ctaggtctaatgcggtgcgctctctctttgacaggcgcggtttggggacaacagcgggg 180
|||||
Db 121 ctaggtctaatgcggtgcgctctctctttgacaggcgcggtttggggacaacagcgggg 180

QY 181 acgagagataaggtgacataccagacagattgtgtgcgcgctgatactctctccg 240
|||||
Db 181 acgagagataaggtgacataccagacagattgtgtgtgcgcgctgatactctctccg 240
```

```
QY 241 acaggaacgcggagctattttaaaagacccctatcgattactttatcttcttggaagct 300
|||||
Db 241 acaggaacgcggagctattttaaaagacccctatcgattactttatcttcttggaagct 300

QY 301 tcttcggagagacaaaaagatgttccctgcctaaagacacaaagggccacacacggaggt 360
|||||
Db 301 tcttcggagagacaaaaagatgttccctgcctaaagacacaaagggccacacacggaggt 360

QY 361 ctgcacagcgacgc 375
|||||
Db 361 ctgcacagcgacgc 375

RESULT 3
AAS09961
ID AAS09961 standard; DNA; 7836 BP.
XX
AC AAS09961;
XX
DT 24-OCT-2001 (first entry)
XX
DE Genomic DNA #1 encoding human Csx/Nkx2.5.
XX
KW Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;
KW therapeutic; heart tissue; gene therapy; human; ds.
XX
OS Homo sapiens.
XX
PN WO200151006-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01511.
XX
PR 14-JAN-2000; 2000US-0176419.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lee IW, Izumo S;
XX
DR WPI; 2001-451809/48.
XX
CC New cardiac specific cell enhancer elements, useful for specifically
CC expressing gene in cardiac cell, as earlier marker of cardiomyocyte
CC induction, e.g. for optimizing cardiomyocyte induction -
XX
PS Disclosure; Fig 4A; 66pp; English.
XX
CC The sequence represents the genomic sequence #1 of human Csx/Nkx2.5. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
SQ Sequence 7836 BP; 2164 A; 1938 C; 1907 G; 1825 T; 2 other;

Query Match 100.0%; Score 375; DB 22; Length 7836;
Best Local Similarity 100.0%; Pred. No. 8.9e-184;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agccccccgacccctacgtgctccgcgcgccttctctccacccctcccgagccctaa 60
|||||
Db 3588 agccccccgacccctacgtgctccgcgcgccttctctccacccctcccgagccctaa 3647

QY 61 agggcgcgggggcccaagcgagggcgctgcgtaccgccgagcggaagggccccagt 120
|||||
Db 3648 agggcgcgggggcccaagcgagggcgctgcgtaccgccgagcggaagggccccagt 3707

QY 121 ctaggtctaatgcggtgcgctctctctttgacaggcgcggtttggggacaacagcgggg 180
```

Db	3708	ctatgtccctaatggcggcgccgtctctcttgaacagcgcggttggagacaacagcgag	3767
QY	181	acqaaataaagtgcataccacgaagattatgaacgcgtgcgtactctctccg	240
Db	3768	acgaqaataaagtgcataccacgaagatttggcgcgcgtgcgtactctctccg	3827
QY	241	acaaqaacgcgcgcgtattttaaagacccctatcgattcttcttcttcggaaagt	300
Db	3828	acaaqaacgcgcgcgtattttaaagacccctatcgattcttcttcttcggaaagt	3887
QY	301	tcttgcgaagagacaaaagatgtccctgcctaaagacacaaagccacacacgcgaagt	360
Db	3888	tcttgcgaagagacaaaagatgtccctgcctaaagacacaaagccacacacgcgaagt	3947
QY	361	ctacacagcgacgc	375
Db	3948	ctgcacagcgacgc	3962

Search completed: July 5, 2002, 06:17:00
Job time: 5398 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 04:46:17 : Search time 80.62 Seconds
(without alignments)
1142.552 Million cell updates/sec

Title: US-09-761-466-1
Perfect score: 375
Sequence: 1 aggcaccccgaccctcacc.....agggtctcacacaggcaacgc 375

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 6

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cqn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cqn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cqn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cqn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cqn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cqn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: July 5, 2002, 06:10:55
Job time: 5078 sec


```
Query Match      100.0%; Score 375; DB 1; Length 375;
Best Local Similarity 100.0%; Pred. No. 2.4e-193;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agggccccgcacccctcatctggctcccgcccttctctccacccctccggaccctctaa 60
Db 1 agggccccgcacccctcatctggctcccgcccttctctccacccctccggaccctctaa 60

Qy 61 agggcgcgcgggcccaagcgagggcgctgcgctgaccccgagcgaaaggcccccagt 120
Db 61 agggcgcgcgggcccaagcgagggcgctgcgctgaccccgagcgaaaggcccccagt 120

Qy 121 ctaggctctaagtcgggtggcgctctctttgacagcgcggtttggggacacacgaggg 180
Db 121 ctaggctctaagtcgggtggcgctctctttgacagcgcggtttggggacacacgaggg 180

Qy 181 acgagagataaggtgacataccagagcagatttggcgcgctgatactctctcccg 240
Db 181 acgagagataaggtgacataccagagcagatttggcgcgctgatactctctcccg 240

Qy 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttctctgaaagct 300
Db 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttctctgaaagct 300

Qy 301 tcttgcggagagacaaaagatttccctgcctaaagacacaaagccacacacgaggggt 360
Db 301 tcttgcggagagacaaaagatttccctgcctaaagacacaaagccacacacgaggggt 360

Qy 361 ctgcacagggcgagcg 375
Db 361 ctgcacagggcgagcg 375

RESULT 2
US-09-761-466-1
; Sequence 1, Application US/09761466
; GENERAL INFORMATION:
; APPLICANT: Lee, Ike W.
; APPLICANT: Izumo, Seigo
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 01948/069002
; CURRENT APPLICATION NUMBER: US/09/761.466
; CURRENT FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-761-466-1
```

```
Query Match      100.0%; Score 375; DB 30; Length 375;
Best Local Similarity 100.0%; Pred. No. 2.4e-193;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agggccccgcacccctcatctggctcccgcccttctctccacccctccggaccctctaa 60
Db 1 agggccccgcacccctcatctggctcccgcccttctctccacccctccggaccctctaa 60

Qy 61 agggcgcgcgggcccaagcgagggcgctgcgctgaccccgagcgaaaggcccccagt 120
Db 61 agggcgcgcgggcccaagcgagggcgctgcgctgaccccgagcgaaaggcccccagt 120

Qy 121 ctaggctctaagtcgggtggcgctctctttgacagcgcggtttggggacacacgaggg 180
Db 121 ctaggctctaagtcgggtggcgctctctttgacagcgcggtttggggacacacgaggg 180
```

```
Qy 181 acgagagataaggtgacataccagagcagatttggcgcgctgatactctctcccg 240
Db 181 acgagagataaggtgacataccagagcagatttggcgcgctgatactctctcccg 240

Qy 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttctctgaaagct 300
Db 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttctctgaaagct 300

Qy 301 tcttgcggagagacaaaagatttccctgcctaaagacacaaagccacacacgaggggt 360
Db 301 tcttgcggagagacaaaagatttccctgcctaaagacacaaagccacacacgaggggt 360

Qy 361 ctgcacagggcgagcg 375
Db 361 ctgcacagggcgagcg 375

RESULT 3
PCT-US01-01511-3
; Sequence 3, Application PC/TUS0101511
; GENERAL INFORMATION:
; APPLICANT: Beth Israel Deaconess Medical Center
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 01948/069W02
; CURRENT APPLICATION NUMBER: PCT/US01/01511
; CURRENT FILING DATE: 2000-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1072
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US01-01511-3
```

```
Query Match      100.0%; Score 375; DB 1; Length 1072;
Best Local Similarity 100.0%; Pred. No. 2.5e-193;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agggccccgcacccctcatctggctcccgcccttctctccacccctccggaccctctaa 60
Db 1 agggccccgcacccctcatctggctcccgcccttctctccacccctccggaccctctaa 60

Qy 61 agggcgcgcgggcccaagcgagggcgctgcgctgaccccgagcgaaaggcccccagt 120
Db 61 agggcgcgcgggcccaagcgagggcgctgcgctgaccccgagcgaaaggcccccagt 120

Qy 121 ctaggctctaagtcgggtggcgctctctttgacagcgcggtttggggacacacgaggg 180
Db 121 ctaggctctaagtcgggtggcgctctctttgacagcgcggtttggggacacacgaggg 180

Qy 181 acgagagataaggtgacataccagagcagatttggcgcgctgatactctctcccg 240
Db 181 acgagagataaggtgacataccagagcagatttggcgcgctgatactctctcccg 240

Qy 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttctctgaaagct 300
Db 241 acaggaacgcggagctatttaaaagaccctatcgattactttatcttctctgaaagct 300

Qy 301 tcttgcggagagacaaaagatttccctgcctaaagacacaaagccacacacgaggggt 360
Db 301 tcttgcggagagacaaaagatttccctgcctaaagacacaaagccacacacgaggggt 360

Qy 361 ctgcacagggcgagcg 375
Db 361 ctgcacagggcgagcg 375

RESULT 4
US-09-761-466-3
```



```
Db 3648 agggcgccggggcccaagccgagggcgctgcgctgaccccgagcggaaggcccccagt 3707
QY 121 ctaggctcctaagcgggcgctcctcttgacagcgcgctttgggacacacagcgggg 180
Db 3708 ctaggctcctaagcgggcgctcctcttgacagcgcgctttgggacacacagcgggg 3767
QY 181 acgagagataaggtgacataccagagcagatttggcgcgctgatactcctctccc 240
Db 3768 acgagagataaggtgacataccagagcagatttggcgcgctgatactcctctccc 3827
QY 241 acaggaaacgcggagctatttaaagaccctatcgatttattcttctctgaaaagct 300
Db 3828 acaggaaacgcggagctatttaaagaccctatcgatttattcttctctgaaaagct 3887
QY 301 tcttgcggagagacaaagatgttcctgcgtaaagacacacaaaggccacacagggagt 360
Db 3888 tcttgcggagagacaaagatgttcctgcgtaaagacacacaaaggccacacagggagt 3947
QY 361 ctgcacagggcgacgc 375
Db 3948 ctgcacagggcgacgc 3962
```

```
RESULT 7
US-60-182-316-1254
; Sequence 1254, Application US/60182316
; GENERAL INFORMATION:
; APPLICANT: Curtis, Anne L.
; APPLICANT: Lagace, Robert E.
; APPLICANT: Klingler, Tod M.
; APPLICANT: Stuve, Laura L.
; FILE REFERENCE: PX-0003 P
; CURRENT APPLICATION NUMBER: US/60/182,316
; CURRENT FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 14,630
; SOFTWARE: PERL Program
; SEQ ID NO 1254
; LENGTH: 338
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: CpG_991027_B15_masked_fa.Contig29958
US-60-182-316-1254
```

```
Query Match 21.3%; Score 80; DB 57; Length 338;
Best Local Similarity 100.0%; Pred. No. 2.3e-32;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 aagctcttcgagagacaaagatgttcctgcgtaaagacacacaaaggccacacacgg 355
Db 10 aagctcttcgagagacaaagatgttcctgcgtaaagacacacaaaggccacacacgg 69

QY 356 agggctctgcacagggcgacgc 375
Db 70 agggctctgcacagggcgacgc 89
```

Search completed: July 5, 2002, 07:32:11
Job time: 9618 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

GM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 05:31:03 : Search time 248.75 Seconds
(without alignments)
2425.605 Million cell updates/sec

Title: us-09-761-466-1
Perfect score: 375
Sequence: 1 agggccccccaccctcgc.....agggtctgcacagggaagc 375

Scoring table: Oligo_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1089956 seqs, 804492379 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents.NA.New.*
1: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
---------------	-------	----------------	--------	-------	-------------

No matches found

Search completed: July 5, 2002, 07:36:39
Job time: 7416 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

GM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 04:44:12 : Search time 2730.19 Seconds
(without alignments)
1853.848 Million cell updates/sec

Title: US-09-761-466-1
Perfect score: 375
Sequence: 1 aagaccccccacaccctcacc.....agggtctcacacaggcgaacac 375

Scoring table: GLI02_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estinu:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	DB	ID
Description					

No matches found

Search completed: July 5, 2002, 05:32:51
Job time: 2919 sec


```
misc_feature 1. 7523
/note="assembly_fragment:01480
fragment_chain:1
clone_end:T7
vector_side:left"
7624. 41360
/note="assembly_fragment:01718
fragment_chain:1"
41461. 47515
/note="assembly_fragment:00204
fragment_chain:1"
47616. 52829
/note="assembly_fragment:02002
fragment_chain:1"
52930. 90493
/note="assembly_fragment:00959
fragment_chain:1"
90594. 109805
/note="assembly_fragment:01522
fragment_chain:1"
109906. 136657
/note="assembly_fragment:01742
fragment_chain:1"
BASE COUNT 33294 a 32671 c 32974 g 37117 t 601 others
ORIGIN
```

```
Query Match 100.0%; Score 51; DB 2; Length 136657;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcctctttaaggcttgatgtcgaactgtcatgtgtacacttaag 51
|||||TTTTTAAAGGCTTGAATGCTGCAACTGTCTATGTGTACACTTAAAG 53506

Db 53556 TGCTCTTTTAAAGGCTTGAATGCTGCAACTGTCTATGTGTACACTTAAAG 53506
```

RESULT 2

```
AC008412
LOCUS AC008412 137545 bp DNA linear HTG 18-JUL-2000
DEFINITION Homo sapiens chromosome 5 clone CTC-281H14, WORKING DRAFT SEQUENCE,
21 ordered pieces.
ACCESSION AC008412
VERSION AC008412.5 GI:9255970
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137545)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
REFERENCE 2 (bases 1 to 137545)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7708838.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 293265
Center clone name: CIT-HSPC_281H14
-----
```

```
Summary Statistics
Consensus quality: 123959 bases at least Q40
Consensus quality: 133398 bases at least Q30
Consensus quality: 134736 bases at least Q20
Estimated insert size: 135340; agarose-fp estimation
Estimated size: 136595; sum-of-contigs estimation
```

```
Quality coverage: 4.95 in Q20 bases; agarose-fp estimation
Quality coverage: 4.91 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
```

```
1 2162: contig of 2162 bp in length
2162: gap of unknown length
2263: contig of 2961 bp in length
5224 5223: gap of unknown length
5324 5224: contig of 9876 bp in length
15200 5324: gap of unknown length
15200 15299: gap of unknown length
22313 15200: contig of 7013 bp in length
22313 22412: gap of unknown length
22413 22313: contig of 14695 bp in length
37108 22413: gap of unknown length
37207 37108: contig of 6656 bp in length
43864 37207: gap of unknown length
43864 43963: gap of unknown length
46965 43864: contig of 3002 bp in length
46966 46965: gap of unknown length
47066 46966: contig of 2412 bp in length
49478 47066: gap of unknown length
49478 49577: gap of unknown length
54523 49478: contig of 4946 bp in length
54524 54523: gap of unknown length
54624 54524: contig of 13791 bp in length
68414 54624: gap of unknown length
68415 68414: gap of unknown length
68515 68415: contig of 4336 bp in length
72851 68515: gap of unknown length
72851 88213: contig of 15263 bp in length
88214 72851: gap of unknown length
88314 88214: contig of 10225 bp in length
88539 88314: gap of unknown length
98639 88539: contig of 9136 bp in length
107774 98639: gap of unknown length
107874 107774: contig of 7271 bp in length
115145 107874: contig of unknown length
115146 115145: gap of unknown length
115246 115146: contig of 3307 bp in length
118552 115246: gap of unknown length
118553 118552: gap of unknown length
128198 118553: contig of 9546 bp in length
128199 128198: gap of unknown length
130123 128199: contig of 1825 bp in length
130124 130123: gap of unknown length
134039 130124: contig of 3815 bp in length
134139 134039: gap of unknown length
136471 134139: contig of 2332 bp in length
136571 136471: gap of unknown length
136571 137545: contig of 975 bp in length.
```

FEATURES source

```
1. 137545
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-281H14"
/clone_lib="Caltech human BAC library C"
BASE COUNT 37277 a 32308 c 31466 g 34494 t 2000 others
ORIGIN
```

```
Query Match 100.0%; Score 51; DB 2; Length 137545;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 tgcctctttaaggcttgatgtcgaactgtcatgtgtacacttaag 51
|||||TTTTTAAAGGCTTGAATGCTGCAACTGTCTATGTGTACACTTAAAG 29522

Db 29472 TGCTCTTTTAAAGGCTTGAATGCTGCAACTGTCTATGTGTACACTTAAAG 29522

RESULT 3
```

AL669821 Homo sapiens chromosome 6 clone XXbac-116A1, linear HTG 31-JAN-2002
LOCUS DEFINITION PROGRESS *** 4 unordered pieces.
ACCESSION AL669821
VERSION AL669821.4 GI:18157151
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Mashreghi-Mohammadi M.
Direct Submission
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Jan 15, 2002 this sequence version replaced gi:18152448.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bpcl16A1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 144176 bases at least Q40
Consensus quality: 144232 bases at least Q30
Consensus quality: 144259 bases at least Q20
Insert size: 144402; sum-of-contigs
Quality coverage: 167909; 14.3% error; agarose-fp
coverage: 9.55x in Q20 bases; sum-of-contigs Quality
coverage: 9.55x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 75928: contig of 75928 bp in length
* 75929 76028: gap of 100 bp
* 76029 116346: contig of 40318 bp in length
* 116347 116446: gap of 100 bp
* 116447 131258: contig of 14812 bp in length
* 13259 131358: gap of 100 bp
* 131359 144702: contig of 13344 bp in length.
*
Location/Qualifiers
1. .144702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXbac-116A1"
/clone_lib="CHORI-501"
1. .75928
/note="assembly_fragment:03023"
fragment_chain:1
clone_end:77
vector_side:left
76029. .116346
/note="assembly_fragment:03493"
fragment_chain:1
116447. .131258
/note="assembly_fragment:03524"
fragment_chain:1
131359. .144702
/note="assembly_fragment:01454"
fragment_chain:1

clone_end:SP4
vector_side:right"
BASE COUNT 41517 a 34170 c 32866 g 36829 t 36% others
ORIGIN
Query Match 100.0%; Score 51; Dh 2; length 144702;
Best Local Similarity 100.0%; Pref. No. 1.3e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tgcctcttttaagaggtttgaatgctctcaactctatctatctacactiaaa 51
Db 4041 TGTCTCTTTTAAGSGGTTGATGTCGCAACTGTATGTCACACTTAAAC 4091
Search completed: July 5, 2002, 06:10:19
Job time: 5.32 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

• Molecule - molecule search, using sw model

Run on: July 5, 2002, 06:17:00 ; Search time 341.99 seconds
(without alignments)

Title:	US-09-761-466-2
Perfect Score:	51

Setting table: cl139_NUC

Searched: 1746436 seqs. 858457221 residues

Word size : 50

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0

1: /STDST/qqqdata/

```
3: /SIDSL/qcdata/
```

```

5: /sids/qcdata/
6: /sids/
7: /sids/qcdata/

```

```

? : /stb51/accdata/
d : /stb51/accdata/

```

```
9: /SDSI/qcdata/
10: /SDSI/qcdata/
```

```
11: /SIDSI/qcdata/
12: /SIDSI/qcdata/
```

```
13: /SIDSL/qcdata/
14: /SIDSL/qcdata/
```

16: /SIDSL/qcdata,

18: /SIDSl/qcadata,

20: /SIDSI/qcdata,

22: /SIPSI/qcdata/

24: /SIDSL/qcdata/

Pred. No. is the number of regression scores greater than or equal to

and is derived by analysis of

[illegible]

51 10.11 22 AM

7	10.0	7836	22	AE
---	------	------	----	----

[illegible]

PD 19-JUL-2001.
XX
XX 16-JAN-2001; 2001WO-US01511.
XX
XX 14-JAN-2000; 2000US-0176419.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX PI Lee IW, Izumo S;
XX
XX WPI; 2001-451809/48.
XX
XX New cardiac specific cell enhancer elements, useful for specifically
PT expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PT induction, e.g. for optimizing cardiomyocyte induction -
XX
XX Claim 1; Fig 5B; 66pp; English.
XX
XX The sequence represents the coding sequence of cardiac enhancer
CC mCsx/Nkx2.5 homology domains A1 + A2 and intervening sequence. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
XX Sequence 1072 BP; 273 A; 290 C; 297 G; 212 T; 0 other;
SQ

Query Match 100.0%; Score 51; DB 22; Length 1072;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tgcctctttaagggttgaaatgtctgcaactgtcatgtgtacacttaaaag 51
|||||
DB 1022 tgcctctttaagggttgaaatgtctgcaactgtcatgtgtacacttaaaag 1072
|||||

RESULT 3
AAS09961
ID AAS09961 standard; DNA; 7836 BP.
XX
XX AC AAS09961;
XX
XX DT 24-OCT-2001 (first entry)
XX
XX DE Genomic DNA #1 encoding human Csx/Nkx2.5.
XX
XX KW Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;
XX KW therapeutic; heart tissue; gene therapy; human; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200151006-A2.
XX
XX PD 19-JUL-2001.
XX
XX 16-JAN-2001; 2001WO-US01511.
XX
XX 14-JAN-2000; 2000US-0176419.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX PI Lee IW, Izumo S;
XX
XX WPI; 2001-451809/48.
XX
XX New cardiac specific cell enhancer elements, useful for specifically
PT expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PT induction, e.g. for optimizing cardiomyocyte induction -
XX
XX Disclosure; Fig 4A; 66pp; English.

XX The sequence represents the genomic sequence #1 of human Csx/Nkx2.5. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
XX Sequence 7836 BP; 2164 A; 1938 C; 1907 G; 1825 T; 2 other;
SQ

Query Match 100.0%; Score 51; DB 22; Length 7836;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tgcctctttaagggttgaaatgtctgcaactgtcatgtgtacacttaaaag 51
|||||
DB 4609 tgcctctttaagggttgaaatgtctgcaactgtcatgtgtacacttaaaag 4659
|||||

Search completed: July 5, 2002, 06:17:00
Job time: 5398 sec

SeqCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

M nucleotide - nucleotide search, using sw model

Run on: July 5, 2002, 06:10:55 : Search time 80.62 Seconds
(without alignments)
155.387 Million cell updates/sec

Title: US-09-761-466-2
Perfect score: 51
Sequence: 1 tgcctctttaaaqcttga.....tgcatatataacttaaa 51

Scoring table: GLCO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum hit seq length: 0
Maximum hit seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cdm2_6/ptodata/2/ina/5A_0.MB.seq.*
2: /cdm2_6/ptodata/2/ina/5B_0.MB.seq.*
3: /cdm2_6/ptodata/2/ina/5A_0.MB.seq.*
4: /cdm2_6/ptodata/2/ina/5B_0.MB.seq.*
5: /cdm2_6/ptodata/2/ina/pt015_0.MB.seq.*
6: /cdm2_6/ptodata/2/ina/backfiles1.seq.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
---------------	-------	----------------	--------	----	----	-------------

No matches found

Search completed: July 5, 2002, 06:10:55
Job time: 5076 sec

GenCore version 4.5
Copyright (c) 1995 - 2000 CompuGen Ltd.

EM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 05:42:51 : Search time 2730.19 seconds
(without alignments)
252.123 Million cell updates/sec

Title: us-09-761-466-2
Perfect score: 51
Sequence: 1 tttctcttttaaaqggttaa.....ttgtcatgtgtacacttaaaa 51

Scoring table: 0.170,NDC
Gapop 60.0 , Gapext 60.0

Searched: 14736207 seqs, 6748477542 residues

Word size : 30

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 9

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estma:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: qb_est1:*
10: qb_est2:*
11: qb_hic:*
12: qb_oss:*
13: em_oss_him:*
14: em_oss_ino:*
15: em_oss_kin:*
16: em_oss_vrt:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
--------	-------	-------	-------	--------	----	----	-------------

87 matches found

Search completed: July 5, 2002, 05:42:51
Job time: 2919 sec

ACCESSION AC008412
VERSION AC008412.5 GI:9255970
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137545)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 137545)
DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7708838.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 293265
Center clone name: CIT-HSPC_281H14

Summary Statistics
Consensus quality: 123959 bases at least Q40
Consensus quality: 133398 bases at least Q30
Consensus quality: 134736 bases at least Q20
Estimated insert size: 135340; agarose-fp estimation
Estimated insert size: 136595; sum-of-contigs estimation
Quality coverage: 4.95 in Q20 bases; agarose-fp estimation
Quality coverage: 4.91 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

Query Match 50.9%; Score 546; DB 2; Length 137545;
Best Local Similarity 99.6%; Pred. No. 1e-308; Indels 4; Gaps 2;
Matches 906; Conservative 0; Mismatches 0;
Qy 1 agggccccgcacccctcatcctggctccgcgcctctctccacccctccgcgcacccctaa 60
Db 28442 AGGCCCCCGCACCCCTCATCTGCTCGCGCCCTTCTCTCCACCCCTCCGCGACCCCTAA 28501
Qy 61 agggcgcgcgccccagccagggcgctgcgcctgaccccgagcgagggccccag 120
Db 28502 AGGGCGCGGGGGCCCAAGCCAGGGCGGCTGCGCCTTACCCGAGCGGAGGCGCCAGT 28561
Qy 121 ctagtctctaagcgggcgctctctcttgacagcgcgctttggggacacacagcggg 180
Db 28562 CTAGGTCTCTAATGCGGGTGGCGTCTCTTTGACAGCGGGCGTGGGGACACAGCGGGG 28621
Qy 181 acgagagataagtgatcacataccagagcagattgtgcgcgcgctgatactctctccg 240
Db 28622 ACGAGAGATAAGTGCATATACAGAGCAGATTTGGTGGCGCGCTGATACTCTCTCCCG 28681
Qy 241 acagagaaacggagcgtattttaaaagacccatcgtatctattcttctct---ggaaa 237
Db 28682 ACAGGAAACGGCGAGCTATTTAAAGACCCCTATCGATTATCTTCTCTGGAGGAA 28741
Qy 298 gctcttcggagagacaaaagatgttcctgccttaagacacaaagccacacacagcgag 357
Db 28742 GCTTCTTCGGGAGACACAAAAGATGTTCCCTGCCCTAAAGACACAGGCCACACAGCGAG 28801
Qy 358 ggtctgcacagcgacgcacaaattcggcgcggggaaagcaaaacacactgacgcttaga 417
Db 28802 GGTCTGCACAGCGACGCAATTCGGCGGGGAAAGCAAAACACACTGACGCTTAGA 28861
Qy 418 gtgcacaaacgtgtgtgttccacagacagctccagagtgcgcgagggagcgctggggcg 477
Db 28862 GTGCACAAACGTTGTGTGTTCACAGACAGCTCCAGAGTGGGCA -GGACGCTGGGGCGG 28920
Qy 478 cgagggcgaccacacagtgatggtctctgtgccttgaaaagtttttttccacgtatg 537
Db 28921 CGAGGGGACCCACACAGTAGTGTGTGTGCTGCTTGGAAAGTTTTTTTTCACCGTATGC 28980
Qy 538 cgtaaaacacgcacacacagagaaagtgtgactgtgcaacttagggcgcgctgtgtaccgt 597
Db 28981 CGTAAACACGACACACACAGAGAAAGTGACTGTGCATTTAGGCGGCTGTGTGTACCCGT 29040
Qy 598 gtctgttttagcgaatttaaaagcacatcagcgcgggccatggctcacgcctgtatccc 657
Db 29041 GTCGTTTTAGCGAATTTAAAGCACATCAGGCGGGCGCCATGGCTCACGCCGTGTAATCCC 29100
Qy 658 aqcacttttagagcgagcgagcgcgccgattcacctaggtcgaggttcacacacagcctg 717
Db 29101 AGCACTTTAGGAGCGCGAGCGGGCGGATCACTTGGGTGGGAGTTCACACACAGCCTG 29160

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165728 bases at least Q40
Consensus quality: 169751 bases at least Q30
Consensus quality: 171461 bases at least Q20
Insert size: 17000; agarose-fp
Insert size: 17337; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1484: contig of 1484 bp in length
* 1485 1584: gap of 100 bp
* 1585 4576: contig of 2992 bp in length
* 4577 4676: gap of 100 bp
* 4677 6987: contig of 2311 bp in length
* 6988 7087: gap of 100 bp
* 7088 11253: contig of 4166 bp in length
* 11254 11353: gap of 100 bp
* 11354 15971: contig of 4618 bp in length
* 15972 16071: gap of 100 bp
* 16072 21310: contig of 5239 bp in length
* 21311 21410: gap of 100 bp
* 21411 32494: contig of 11084 bp in length
* 32495 32594: gap of 100 bp
* 32595 47895: contig of 15301 bp in length
* 47896 47995: gap of 100 bp
* 47996 64361: contig of 16366 bp in length
* 64362 64461: gap of 100 bp
* 64462 81134: contig of 16673 bp in length
* 81135 81234: gap of 100 bp
* 81235 105677: contig of 24443 bp in length
* 105678 105777: gap of 100 bp
* 105778 135680: contig of 29903 bp in length
* 135681 135780: gap of 100 bp
* 135781 174537: contig of 38757 bp in length.

FEATURES

source
1. .174537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-805N8"
/clone_lib="RPC1-11 Human Male BAC"
1. .1484
/note="assembly_fragment"
1585. 4576
/note="assembly_fragment"
4677. 6987
/note="assembly_fragment"
clone_end:SP6
vector_side:left
7088. 11253
/note="assembly_fragment"
11354. 15971
/note="assembly_fragment"
16072. 21310
/note="assembly_fragment"
clone_end:T7
vector_side:left
21411. 32494
/note="assembly_fragment"
32595. 47895
/note="assembly_fragment"
47996. 64361
/note="assembly_fragment"

misc_feature 64462..81134

/note="assembly_fragment"

misc_feature 81235..105677

/note="assembly_fragment"

misc_feature 105778..135680

/note="assembly_fragment"

misc_feature 135781..174537

/note="assembly_fragment"

BASE COUNT 55112 a 33702 c 33796 g 50727 t 1200 others

ORIGIN

Query Match 9.1%; Score 98; DB 2: Length 174537;

Best Local Similarity 100.0%; Pred. NO. 1.2e-45;

Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 tcccagctactcgggaggtgagcgaggaatcgcttgaacccggagcgagggtgc 848
|||||

Db 86180 TCCAGCTACTCGGAGGCTGAGGCGAGGAGATCGCTTGAAACCGGCGGAGGTTGC 86239
|||||

QY 849 agtgagccgagatcacaccactgcactccagcctgggc 886
|||||

Db 86240 AGTGAGCCGAGATCACACCCTCCACTCCAGCCTGGGC 86277
|||||

RESULT 6

AP000871/c

LOCUS AP000871 189492 bp DNA linear HTG 22-NOV-2000

DEFINITION Homo sapiens chromosome 11 clone RP11-709J7 map 11q23, WORKING

DRIFT SEQUENCE, 19 unordered pieces.

ACCESSION AP000871

VERSION AP000871.3 GI:11320823

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens DNA, clone:RP11-709J7.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 189492)

1 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Homo sapiens 189,492 genomic DNA of 11q23

Published Only in DataBase (1999) In press

2 (bases 1 to 189492)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission

Submitted (13-DEC-1999) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,

Tel:81-45-503-9111, Fax:81-45-503-9170)

On Nov 22, 2000 this sequence version replaced gi:8119019.

COMMENT

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft11

Center clone name: RP11-709J7

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 184605 bases at least Q40

Consensus quality: 186253 bases at least Q30

Consensus quality: 187002 bases at least Q20

Insert size: 187692; sum-of-contigs

Quality coverage: 8.38x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
19 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the


```
JOURNAL      Published Only in DataBase (2000) In press
REFERENCE    2 (bases 1 to 211344)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
              (E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
              Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      On Mar 16, 2001 this sequence version replaced gi:11559295.
FEATURES     location/Qualifiers
             1..211344
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="11"
             /map="11q"
             /clone="RP11-801G16"
BASE COUNT   61435 a 41254 c 42802 g 65853 t
ORIGIN

Query Match          9.1%; Score 98; DB 9; Length 211344;
Best Local Similarity 100.0%; Pred. No. 1.2e-45;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 789 tccagctactcggagcgtgagcagcaggaatcgttgaacccggagcgaggttgc 848
      |||||||
Db 68130 TCCAGCTACTCGGAGGCTGAGCGAGGAGGAATCTTGAACCGGAGCGAGGTTGC 68189

Oy 849 agtgagcagatcacacactgcactcagcctggcg 886
      |||||||
Db 68190 AGTGAGCGGAGATCACACCTGCACCTCCAGCGCTGGC 68227

RESULT      8
LOCUS       AC016765/c
DEFINITION Homo sapiens chromosome 11 clone RP11-555F1, WORKING DRAFT
ACCESSION   AC016765
VERSION     AC016765.2 GI:7232204
KEYWORDS    HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
            2 (bases 1 to 181047)
            Waterston,R.H.
            Direct Submission
            Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Mar 13, 2000 this sequence version replaced gi:6524271.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0555F01
----- Summary Statistics -----
Sequencing vector: M13; 81%
Sequencing vector: plasmid; 19%
Chemistry: Dye-terminator Big Dye; 19% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 166484 bases at least Q40
Consensus quality: 171727 bases at least Q30

Consensus quality: 174384 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 179147; sum-of-contigs
Quality coverage: 3.63 in Q20 bases; agarose-fp
Quality coverage: 3.93 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1833: contig of 1833 bp in length
* 1834: gap of unknown length
* 1934: contig of 1937 bp in length
* 3871: gap of unknown length
* 3971: contig of 2916 bp in length
* 6887: gap of unknown length
* 6987: contig of 3066 bp in length
* 10052: gap of unknown length
* 10153: contig of 4031 bp in length
* 14184: gap of unknown length
* 14284: contig of 2980 bp in length
* 17264: gap of unknown length
* 17364: contig of 5825 bp in length
* 23189: gap of unknown length
* 23289: contig of 4717 bp in length
* 28006: gap of unknown length
* 28106: contig of 4500 bp in length
* 32706: gap of unknown length
* 32706: contig of 5504 bp in length
* 38210: gap of unknown length
* 38310: contig of 6480 bp in length
* 44790: gap of unknown length
* 44890: contig of 6670 bp in length
* 51560: gap of unknown length
* 51560: contig of 7526 bp in length
* 59186: gap of unknown length
* 59286: contig of 8041 bp in length
* 67327: gap of unknown length
* 67427: contig of 6165 bp in length
* 73592: gap of unknown length
* 73692: contig of 8772 bp in length
* 82464: gap of unknown length
* 82564: contig of 8149 bp in length
* 90713: gap of unknown length
* 90813: contig of 19225 bp in length
* 110038: gap of unknown length
* 110138: contig of 27128 bp in length
* 137266: gap of unknown length
* 137366: contig of 43682 bp in length.

Location/Qualifiers
1..181047
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-555F1"
/contig="assembly_name:Contig13"
/contig="assembly_name:Contig14"
/contig="assembly_name:Contig15"
vector_side:right"
6987..10052
/contig="assembly_name:Contig16"
10153..14183
/contig="assembly_name:Contig17"
14284..17263
/contig="assembly_name:Contig18"

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
```



```
* 23774 27287: contig of 3514 bp in length
* 27288 27387: gap of 100 bp
* 27388 33448: contig of 6061 bp in length
* 33448 33548: gap of 100 bp
* 33549 39008: contig of 5460 bp in length
* 39009 39108: gap of 100 bp
* 39109 44224: contig of 5116 bp in length
* 44225 44324: gap of 100 bp
* 44325 50232: contig of 5908 bp in length
* 50233 50332: gap of 100 bp
* 50333 58422: contig of 8090 bp in length
* 58423 58522: gap of 100 bp
* 58523 64440: contig of 5918 bp in length
* 64441 64540: gap of 100 bp
* 64541 72727: contig of 8187 bp in length
* 72728 72827: gap of 100 bp
* 72828 81027: contig of 8200 bp in length
* 81028 81127: gap of 100 bp
* 81128 91377: contig of 10250 bp in length
* 91378 91477: gap of 100 bp
* 91478 102460: contig of 10983 bp in length
* 102461 102560: gap of 100 bp
* 102561 117037: contig of 14477 bp in length
* 117038 117137: gap of 100 bp
* 117138 129181: contig of 12044 bp in length
* 129182 129281: gap of 100 bp
* 129282 149602: contig of 20321 bp in length
* 149603 149702: gap of 100 bp
* 149703 178508: contig of 28806 bp in length.
Location/Qualifiers
1..178508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-748N13"
1..1348
/note="assembly_fragment"
1449..2859
/note="assembly_fragment"
2960..5254
/note="assembly_fragment"
5355..8788
/note="assembly_fragment"
8889..11519
/note="assembly_fragment"
clone_end:17
vector_side:left"
11620..15398
/note="assembly_fragment"
15499..19979
/note="assembly_fragment"
20080..23673
/note="assembly_fragment"
23774..27287
/note="assembly_fragment"
27388..33448
/note="assembly_fragment"
33549..39008
/note="assembly_fragment"
39109..44224
/note="assembly_fragment"
44325..50232
/note="assembly_fragment"
50333..58422
/note="assembly_fragment"
58523..64440
/note="assembly_fragment"
64541..72727
/note="assembly_fragment"
72828..81027
/note="assembly_fragment"
```

FEATURES
source

```
misc_feature 81128..91377
/note="assembly_fragment"
misc_feature 91478..102460
/note="assembly_fragment"
misc_feature 102561..117037
/note="assembly_fragment"
misc_feature 117138..129181
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
129282..149602
/note="assembly_fragment"
149703..178508
/note="assembly_fragment"
BASE COUNT 45506 a 42854 c 42878 g 45067 t 2203 others
ORIGIN

Query Match 8.3%; Score 89; DB 2; Length 178508;
Best Local Similarity 100.0%; Pred. No. 2.4e-40;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 802 ggaggctgaggcaggagaatcgcttgaacccggaggcggaggttcagtgagccgagat 861
|||||
Db 96010 GGAGGCTGAGCGAGGAGAAATCGCTTGAACCGGAGCGGAGGTTGCAGTGAGCCGAGAT 95951
|||||

QY 862 cacaccactgcactccagcctggcgaca 890
|||||

Db 95950 CACACCACCTGCACCTCCAGCCTGGCGACA 95922
|||||

RESULT 12
AC010984
LOCUS AC010984 191540 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-510C1 from 2, complete sequence.
ACCESSION AC010984
VERSION AC010984.7 GI:15668160
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 191540)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 191540)
AUTHORS Du,F., Maupin,R. and Hawkins,M.
TITLE The sequence of Homo sapiens BAC clone RP11-510C1
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 191540)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 191540)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 191540)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 19, 2001 this sequence version replaced gi:14349340.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
```


Contact: sapiens@atlas.wustl.edu
----- Summary Statistics
Center project name: R:NM0510c91

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping cDNA sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John B. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCL11 human BAC library was made from the blood of one male donor, as described by Osomawa, K., Woon, P.-Y., Zhao, R., Frengen, E., Tateno, M., Cattanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECLINK: pBAC63.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-37H17, 2000 bp overlap; the clone sequenced to the right is AC023040. Actual start of this clone is at base position 29090 of RP11-109KL2.

FEATURES

source

location/qualifiers

1..191540

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/map ">"

/clone="RP11-510c1"

/clone_lib="RPCL11"

55..1455

/note="match to EST AM632464 (NID:a2555878) np86e08.s1"

418..1940

/note="similar to Homo sapiens EST AV742195

(NID:a10849740)"

761..1010

/note="match to EST BF455538 (NID:a1144612)"

795..1415

/note="match to EST BF167120 (NID:a10889561)"

1028..1189

/note="match to EST BF350015 (NID:a9261868) ht07ncl.x1"

1028..1385

/note="match to EST AL041002 (NID:a9442428)"

1034..1504

/note="match to EST AL44427 (NID:a4270658) te4crl1.x1"

1469..1391

/rpt_family="AT_10c1"

1426..12966

/note="similar to Sus scrofa EST R1185646 (NID:a3465997)"

1426..1763

/note="similar to Homo sapiens EST A1826796 (NID:a35447467)

w856d04.x1"

1426..1763

/note="similar to Homo sapiens EST A1805283 (NID:a5491849)

to2408.x1"

1426..1763

/note="similar to Homo sapiens EST BP027338

misc_feature

(NID:a107565c1)"

/note="similar to Homo sapiens EST A1762617 (NID:a177619e)

wh9ld06.x1"

1557..11003

/note="similar to Bos taurus EST BF60486 (NID:a17066f4)"

2101..13428

/note="cpd_island (x=0.000000, y=0.000000, #cpds=06)"

2142..2345

/note="match to EST AC141414 (NID:a141418)"

2511..12544

/rpt_family="A..."

2651..13261

/rpt_family="A1..."

2936..12962

/rpt_family="(CAA)..."

3298..13616

/note="match to EST BF344635 (NID:a9606767)"

3659..14286

/note="match to EST BF442299 (NID:a1194994)"

4008..13667

/rpt_family="L1"

4724..14622

/note="match to EST BF48699 (NID:a12465974)"

4777..15075

/note="similar to Homo sapiens EST BF750345

(NID:a12077021)"

4940..15074

/note="similar to Homo sapiens EST BF567017

(NID:a9810781)"

4943..15074

/note="match to EST BF591155 (NID:a12477534)"

4950..15174

/note="match to EST BF42699 (NID:a12465974)"

4991..15165

/note="match to EST AW206302 (NID:a9606767)"

5284..15483

/note="similar to Homo sapiens EST AM54287 (NID:a2590441)

np86d08.s1"

5442..15498

/note="similar to Homo sapiens EST A1799417 (NID:a9606767)

113404.x1"

/note="match to EST BF44444 (NID:a12465974)"

6800..15897

/note="similar to Homo sapiens EST A1799417 (NID:a9606767)

113404.x1"

/note="match to EST BF960155 (NID:a12477534)"

5847..15897

/note="match to EST AM64969 (NID:a2606115) pr24a1.11"

5850..15897

/note="match to EST AA44721 (NID:a12465974) os11b.2.81"

5856..15897

/note="match to EST BF47731 (NID:a12465974)"

6129..16229

/note="match to EST BF455538 (NID:a1144612)"

6482..16670

/rpt_family="MBP_type"

7029..17462

repeat_region

7194..17672

/rpt_family="L1..."

7445..17682

/note="similar to Homo sapiens EST AV12455

(NID:a10447470)"

7673..17812

/rpt_family="A1..."

7609..17696

/note="match to EST AM7144 (NID:a271958) ad011.0.11"

7613..17699

/rpt_family="L1"

```

repeat_region 7898..7993 /rpt_family="L1"
repeat_region 8000..8419 /rpt_family="L1"
repeat_region 8409..8448 /rpt_family="(T)n"
repeat_region 8420..8735 /rpt_family="Alu"
misc_feature 8619..8651 /note="match to EST AA704040 (NID:g2713958) ag81f10.r1"
repeat_region 8736..9064 /rpt_family="L1"
repeat_region 9099..9183 /rpt_family="MIR"

Query Match 8.3% Score 89; DB 9; Length 191540;
Best Local Similarity 100.0%; Pred. No. 2.4e-40;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 802 ggagcgtgagcagaggaatccttgaaacccggagcgaggttcagtgagccgagat 861
|||||
Db 41415 GGAGGCTGAGGCAGAGAAATCGCTTGAAACCGGAGCGGAGGTTCAGTCAGCGGAGAT 41474
|||||

QY 862 cacaccactgacactcagcctg9gcgaca 890
|||||
Db 41475 CACACCACCTGCACTCCAGCCTGGCGGACA 41503
|||||

RESULT 13
AC025197 151865 bp DNA linear HTG 20-SEP-2000
LOCUS Homo sapiens chromosome 8 clone RP11-209E9 map 8, WORKING DRAFT
DEFINITION AC025197 SEQUENCE, 14 unordered pieces.
ACCESSION AC025197.2 GI:10198530
VERSION AC025197.2
KEYWORDS HTG; HTGS-PHASEI; HTGS-DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-209E9
Unpublished
2 (bases 1 to 151865)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boquslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castler,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lenoczky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,K., Schauer,S., Severly,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:7188875.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

FEATURES
source
Location/Qualifiers
1..151865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-209E9"
/clone_lib="RPCI-11 Human Male BAC"
1..9341
/note="assembly_fragment
clone_end:SP6
vector_side:left"
9442..11373
/note="assembly_fragment"
11474..47057
/note="assembly_fragment"
47158..49227
/note="assembly_fragment"

misc_feature
misc_feature
misc_feature
misc_feature

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7490
Center clone name: 209_E9
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 146689 bases at least Q40
Consensus quality: 149124 bases at least Q30
Consensus quality: 149969 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 150565; sum-of-contigs
Quality coverage: 5.9 in Q20 bases; agarose-fp
Quality coverage: 5.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 9342 9441: gap of 100 bp
* 9442 11373: contig of 1932 bp in length
* 11374 11473: gap of 100 bp
* 11474 47057: contig of 35584 bp in length
* 47058 47157: gap of 100 bp
* 47158 49227: contig of 2070 bp in length
* 49228 49327: gap of 100 bp
* 49328 54639: contig of 5312 bp in length
* 54640 54739: gap of 100 bp
* 54740 59242: contig of 4503 bp in length
* 59243 59342: gap of 100 bp
* 59343 65284: contig of 5942 bp in length
* 65285 65384: gap of 100 bp
* 65385 70923: contig of 5539 bp in length
* 70924 71023: gap of 100 bp
* 71024 78085: contig of 7062 bp in length
* 78086 78185: gap of 100 bp
* 78186 87419: contig of 9234 bp in length
* 87420 87519: gap of 100 bp
* 87520 97716: contig of 10197 bp in length
* 97717 97816: gap of 100 bp
* 97817 120620: contig of 22804 bp in length
* 120621 120720: gap of 100 bp
* 120721 150210: contig of 29490 bp in length
* 150211 150310: gap of 100 bp
* 150311 151865: contig of 1555 bp in length.
Location/Qualifiers
1..151865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-209E9"
/clone_lib="RPCI-11 Human Male BAC"
1..9341
/note="assembly_fragment
clone_end:SP6
vector_side:left"
9442..11373
/note="assembly_fragment"
11474..47057
/note="assembly_fragment"
47158..49227
/note="assembly_fragment"

was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>
RP11-535M15 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-535M15 it may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-330M2 is at 140528 in this sequence. The true right end of clone RP11-392G7 is at 2000 in this sequence.

FEATURES

source	Location/Qualifiers
1	.142527
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="9"
	/clone="RP11-535M15"
	/clone_lib="RPCI-11.2"
BASE COUNT	40707 a 29669 c 30360 g 41791 t
ORIGIN	

Query Match 7.9%; Score 85; DB 9; Length 142527;
Best Local Similarity 100.0%; Pred. No. 5.2e-38;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 802 ggagcctgagcagagaatcgcttgaaacccggagcgaggttgacgtgagccagat 861
|||||
Db 7322 GGAGCCTGAGCGAGGAGATCGCTTGAACCCGGGAGCGGAGTTGCAGTGAGCCGAGAT 7263
|||||

Qy 862 cacaccactgcactccagcctgggc 886
|||||

Db 7262 CACACCACTGCACCTCCAGCCTGGGC 7238

Search completed: July 5, 2002, 06:24:06
Job time: 5959 sec


```

; FILING DATE: October 17, 1994
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1656
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-465-981-2

```

```

Query Match 5.7%; Score 61; DB 2; Length 1656;
Best Local Similarity 100.0%; Pred. No. 1.7e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 802 qaagcttaagcagaaatcacttgaacccgggaagcagaaatgagcagaaat 861
|||||
DB 369 GGAGGCTGAGGCGAGCAATCGCTTGAACTGGGAGCGGAGGCTTGCAAGTGGCGGAGAT 310
|||||

QY 862 c 862
+
DB 309 C 309

```

```

RESULT 5
PCI-US93-11915-2/c
; Sequence 2; Application PC/TUS9311915
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Ace, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCI/US93/11915
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1656
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
PCI-US93-11915-2

Query Match 5.7%; Score 61; DB 5; Length 1656;
Best Local Similarity 100.0%; Pred. No. 1.7e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 802 qaagcttaagcagaaatcacttgaacccgggaagcagaaatgagcagaaat 861
|||||
DB 369 GGAGGCTGAGGCGAGCAATCGCTTGAACTGGGAGCGGAGGCTTGCAAGTGGCGGAGAT 310
|||||

QY 862 c 862
+
DB 309 C 309

```

```

RESULT 6
US-08-324-465-5/c
; Sequence 5; Application US/08324465
; Patent No. 5565314
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Ace, Miyako
; TITLE OF INVENTION: GENE TRANSCRIPTION AND
; TITLE OF INVENTION: IONIZING RADIATION: METHODS
; TITLE OF INVENTION: AND COMPOSITIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,465
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/999,742
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-324-465-5

```

```

Query Match 5.7%; Score 61; DB 1; Length 1725;
Best Local Similarity 100.0%; Pred. No. 1.7e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 802 qaagcttaagcagaaatcacttgaacccgggaagcagaaatgagcagaaat 861
|||||
DB 369 GGAGGCTGAGGCGAGCAATCGCTTGAACTGGGAGCGGAGGCTTGCAAGTGGCGGAGAT 310
|||||

```

Qy 862 c 862
Db 309 C 309

RESULT 7

US-08-465-981-5/c
; Sequence 5, Application US/08465981
; Patent No. 5874415
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,981
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/324,465
; FILING DATE: October 17, 1994
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

US-08-465-981-5

Query Match 5.7%; Score 61; DB 2; Length 1725;
Best Local Similarity 100.0%; Pred. No. 1.7e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 802 ggaggtcaggcaggagaaatcgcttgaacccggaggcggaggttgacgtgagccgagat 861
|||||
Db 369 GGAGGCTGAGCAGGAGAAATCGCTTGAACCCGGAGCGGAGGTTGCAGTGAGCCGAGAT 310

Qy 862 c 862
Db 309 C 309

RESULT 8

PCT-US93-11915-5/c
; Sequence 5, Application PC/TUS9311915
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING

; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11915
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US93-11915-5

Query Match 5.7%; Score 61; DB 5; Length 1725;
Best Local Similarity 100.0%; Pred. No. 1.7e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 802 ggaggtcaggcaggagaaatcgcttgaacccggaggcggaggttgacgtgagccgagat 861
|||||
Db 369 GGAGGCTGAGCAGGAGAAATCGCTTGAACCCGGAGCGGAGGTTGCAGTGAGCCGAGAT 310

Qy 862 c 862
Db 309 C 309

RESULT 9

US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0

SEQUENCE INFORMATION
: SEQUENCE CHARACTERISTICS:
: LENGTH: 72444
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURES:
: NAME/KEY: modified base
: LOCATION: all n positions
: OTHER INFORMATION: n-a, c, q, or t
US-09-268-492 7

Query Match: 5.7%; Score 61; DB 4; Length 72404;
Best Local Similarity 100.0%; Pred. No. 1.5e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 801 aaagagctgaggcaggaataatccttgaacccaggaagcagagattacatgaacccgaaa 860
|||||
DB 15892 agagagctgaggcaggaataatccttgaacccaggaagcagagattgcaatgaacccgaaa 15951
QY 861 t 861
DB 15952 t 15952

RESULT 10
US-08-618-100B-3
: Sequence 3, Application US/08618100B
: Patent No. 6068976
: GENERAL INFORMATION:
: APPLICANT: Briggs, Michael R.
: APPLICANT: Auerk, Johan
: APPLICANT: de Vos, Piet
: APPLICANT: Steels, Bart
: APPLICANT: Croston, Glenn E.
: APPLICANT: Miller, Stephen G.
: TITLE OF INVENTION: MODULATORS OF GENE AND
: TITLE OF INVENTION: SCREENING METHODS THEREFOR
: NUMBER OF SEQUENCES: 48
: COMPRESSION ADDRESS:
: ADDRESS: Lyon & Lyon
: STREET: 615 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: Storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/618,100B
: FILING DATE: March 19, 1996
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/558,588
: FILING DATE: October 30, 1995
: APPLICATION NUMBER: 08/510,584
: FILING DATE: August 2, 1995
: APPLICATION NUMBER: 08/418,036
: FILING DATE: April 5, 1995
: APPLICATION NUMBER: 08/408,534
: FILING DATE: March 20, 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,427
: REFERENCE/DOCKET NUMBER: 219/075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 555-0440
: TELEX: 67-3510

SEQUENCE INFORMATION
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10684 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: lna (nucleic)
: DESCRIPTION: Sequence between exons 1 and exon 2
: Patent No. 6068976
US-08-618-100B-4

Query Match: 5.4%; Score 54; DB 4; Length 10684;
Best Local Similarity 100.0%; Pred. No. 4.5e-14;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 802 aaagctgaacccaataatcgtttaaaccccaataacagaaatttgaatgaacccaa 859
|||||
DB 4657 GGAAGCTGAGCGAGGAGAAAGCTTGAAAGGAGAAAGGAGAAAGTTTGAATGAGCGAG 4714

RESULT 11
US-09-328-111-741/0
: Sequence 741, Application US/09328111
: Patent No. 6262-33
: GENERAL INFORMATION:
: APPLICANT: Endoche, Willson B.
: APPLICANT: Steinmann, Kathleen E.
: APPLICANT: Astle, Jon B.
: APPLICANT: Barress, Christy Beth C.
: APPLICANT: Bushnell, Steven E.
: APPLICANT: Carroll III, Eddie
: APPLICANT: Catino, Theodore J.
: APPLICANT: Dertli, Adam
: APPLICANT: Ford, Donna M.
: APPLICANT: Lewis, Marcia E.
: APPLICANT: Monahan, John E.
: APPLICANT: Schlessel, Robert
: TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
: TITLE OF INVENTION: PRODOTS
: FILE REFERENCE: CDB-257 (US)
: CURRENT APPLICATION NUMBER: US/09/328,111
: CURRENT FILING DATE: 1999-06-08
: EARLIER APPLICATION NUMBER: US 60/388,801
: EARLIER FILING DATE: 1998-06-10
: NUMBER OF SEQ ID NOS: 850
: SOFTWARE: FastSeq for Windows Version 2.0
: SEQ ID NO 741
: LENGTH: 679
: TYPE: lna
: ORGANISM: Homo sapiens
: FEATURES:
: NAME/KEY: misc feature
: LOCATION: (1)...(579)
: OTHER INFORMATION: n-a, A, C or G
US-09-328-111-741

Query Match: 5.0%; Score 54; DB 4; Length 848;
Best Local Similarity 100.0%; Pred. No. 4.5e-17;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 795 ctactcandaataatgaaccccaataatcgtttaaaccccaataacagaaatttgaatgaacccaa 848
|||||
DB 213 GATATGGAGGCTGAGGTAAGAAATCTTGAAGAGGAGAAAGTTTGAATGAGCGAG 163

RESULT 12
US-09-242-948-1/0
: Sequence 3, Application US/09242948
: Patent No. 6252057
: GENERAL INFORMATION:
: APPLICANT: Brady, Matthew J.

; Printen, John A
; Saltiel, Alan R
; Warner-Lambert Company,
; (Outside USA)
; TITLE OF INVENTION: Protein Targeting to Glycogen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 201 Tabor Road
; CITY: Morris Plains
; STATE: NJ
; COUNTRY: US
; ZIP: 07950
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/242,948
; FILING DATE: 25-Feb-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,107
; FILING DATE: 30-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ashbrook, Charles W
; REFERENCE/DOCKET NUMBER: 5485-01-CA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313 996-5215
; TELEFAX: 313 996-1553
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4238..5176
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-242-948-3

Query Match 5.0%; Score 54; DB 4; Length 5789;
Best Local Similarity 100.0%; Pred. No. 4.2e-17;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 802 ggagcgtgagcagagagaatcgcttgaaacccggaggcggaggttgagtgagc 855
|||||
Db 311 GGAGCGTAGGCAGGAGGAATCGCTTGAACCCGGAGCGGAGGTTCAGTGAGC 258

RESULT 13
US-09-210-748A-3/C
; Sequence 3, Application US/09210748A
; Patent No. 6335156
; GENERAL INFORMATION:
; APPLICANT: Hermeking, Heiko
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: 14-3-3 SIGMA ARREST THE CELL CYCLE
; FILE REFERENCE: 1107.77810
; CURRENT APPLICATION NUMBER: US/09/210,748A
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/069,416
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 7680
; TYPE: DNA

; ORGANISM: Homo sapiens
US-09-210-748A-3
Query Match 5.0%; Score 54; DB 4; Length 7680;
Best Local Similarity 100.0%; Pred. No. 4.2e-17;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 789 tcccagctactcggagcgtgagcagagagaatcgcttgaaacccggaggcggga 842
|||||
Db 6086 TCCAGCTACTCGGAGGCTGAGGCAGGAGATCGCTTGAACCCGGAGCGGA 6033
RESULT 14
US-09-247-155-141
; Sequence 141, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bouqueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 141
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..447
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 4..147
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.6999980926514
; OTHER INFORMATION: seq LLLFFKLLVVG/VG
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 858..863
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 880..891
US-09-247-155-141
Query Match 4.9%; Score 52; DB 4; Length 891;
Best Local Similarity 100.0%; Pred. No. 4.2e-16;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 788 atcccagctactcggagcgtgagcagagagaatcgcttgaaacccggaggc 839
|||||
Db 714 atcccagctactcggagcgtgagcagagagaatcgcttgaaacccggaggc 765
RESULT 15
US-08-973-544-1
; Sequence 1, Application US/08973544
; Patent No. 6338950
; GENERAL INFORMATION:
; APPLICANT: WEISS, Elisabeth
; TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1

```
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"
166 a 121 c 131 g 118 t

Query Match 7.9%; Score 85; DB 12; Length 536;
Best Local Similarity 100.0%; Pred. No. 7.4e-32;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 tgaagcagagaatcgcttgaaacccgagcgaggttgagtgagcgagatcacacc 867
Db 108 TGAGCAGGAGAAATCGCTTGAAACCGGAGCGGAGGTTCAGTGCAGCGAGATCACACC 167
Qy 868 actgcactccagctggggcgacaag 892
Db 168 ACTGCACCTCCAGCTGGGGGACAG 192

RESULT 2
N76766/c
LOCUS
DEFINITION
N76766 494 bp mRNA linear EST 28-JAN-1997
IMAGE:245672 5' similar to contains Alu repetitive element;; mRNA
sequence.
N76766
EST.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 494)
AUTHORS
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissee,S., Dietrich,N., DuBuque,T., Favellio,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasakis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 845 Std Error: 0.00
Seq primer: reverse ET
High quality sequence stop: 374.
Location/Qualifiers
1..494
/organism="Homo sapiens"
/db_xref="GDB:3794918"
/db_xref="taxon:9606"
/clone="IMAGE:245672"
/clone_lib="Soares fetal liver spleen 1NFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5', AACTGGAGAAATTAATTAAGATCTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
108 a 142 c 119 g 121 t 4 others

BASE COUNT
Query Match 7.6%; Score 81; DB 9; Length 423;
Best Local Similarity 100.0%; Pred. No. 8.2e-30;

ORIGIN
```

```
Query Match 7.6%; Score 82; DB 10; Length 494;
Best Local Similarity 100.0%; Pred. No. 2.5e-30;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 801 gggagctgagcgagcgaggaatcgcttgaaacccgggagcgaggttgagtgagcgagga 860
Db 107 GGGAGCTGAGCGGAGGAGAAATCGCTTGAAACCGGAGCGGAGGTTCAGTGCAGCGCAGA 48
Qy 861 tcacaccactgcactcagcct 882
Db 47 TCACACCACTGCACCTCAGCCT 26

RESULT 3
AA224995/c
LOCUS
DEFINITION
AA224995 423 bp mRNA linear EST 21-AUG-1997
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION
AA224995
VERSION
AA224995.1 GI:1846286
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 423)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
, Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 600 Std Error: 0.00
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 353.
Location/Qualifiers
1..423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1010024"
/clone_lib="NCI-CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."
110 a 109 c 92 g 108 t 4 others

BASE COUNT
Query Match 7.6%; Score 81; DB 9; Length 423;
Best Local Similarity 100.0%; Pred. No. 8.2e-30;

ORIGIN
```


TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Other ESTs: bb1le04.y1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -400P from Gibco
High quality sequence stop: 446.

FEATURES source

1..463
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2959134"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
88 a 147 c 109 g 119 t

BASE COUNT ORIGIN

Query Match 7.0%; Score 75; DB 10; Length 463;
Best Local Similarity 100.0%; Pred. No. 8.9e-27;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 788 atcccgactactcggaggctgagcgaggaatcgcttgacccggagcgaggttg 847
|||||
Db 285 ATCCAGCTACTCGGAGGCTGAGCAGGAGAAATCGCTTGACCCGGAGCGAGGTG 226
|||||

Qy 848 cagtgcgcgagatc 862
|||||
Db 225 CAGTGAGCGGAGATC 211

RESULT 7 AW957414

LOCUS AW957414 526 bp mRNA linear EST 01-JUN-2000
DEFINITION EST369604 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AW957414
VERSION AW957414.1 GI:8147217
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 526)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.

Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)

JOURNAL

Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 107
Seq primer: Reverse.

FEATURES source

Location/Qualifiers
1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSkm" 117 t 1 others
150 a 108 c 150 g

BASE COUNT ORIGIN

Query Match 7.0%; Score 75; DB 9; Length 526;
Best Local Similarity 100.0%; Pred. No. 8.7e-27;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 788 atcccgactactcggaggctgagcgaggaatcgcttgacccggagcgaggttg 847
|||||
Db 260 ATCCAGCTACTCGGAGGCTGAGCAGGAGAAATCGCTTGACCCGGAGCGAGGTG 319
|||||

Qy 848 cagtgcgcgagatc 862
|||||
Db 320 CAGTGAGCGGAGATC 334

RESULT 8 AA527961

LOCUS AA527961 569 bp mRNA linear EST 05-AUG-1997
DEFINITION nh30e09.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:953896
similar to contains Alu repetitive element; contains element MER29
MER29 repetitive element ;, mRNA sequence.

ACCESSION AA527961
VERSION AA527961.1 GI:2270030
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 569)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

REFERENCE AUTHORS TITLE JOURNAL COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
, Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 564 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 387.

FEATURES source

Location/Qualifiers
1..569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:953896"
/clone_lib="NCI_CGAP_Pr3"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
000 microdissected cells histologically-determined to be
fully malignant prostate cancer cells. Double-stranded
cDNA was ligated to EcoRI adaptors, 5 cycles of PCR
applied to the cDNA with an adaptor-specific primer, and
the resulting PCR product subcloned into pAMP10 by the
UDG-cloning method (Life Technologies). Average insert
size is 600 bp. NOTE: Not directionally cloned. This

BASE COUNT 163 a 133 c 162 g 106 t
ORIGIN library was constructed by David Krizman."

Query Match 7.0%; Score 75; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 8,6e-27;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 788 tcccagctactcggaggctgaacgaagaatcgccttgaacccggagcgaacgaatc 847
|||||
Db 351 ATCCGAGCTACTCGGAGGCTGAGCGACGAGAAATCGCTTGAACCGGCGAGCGAGGTTG 410
|||||

QY 848 caatgaacgaatc 862
|||||

Db 411 CAGTGGCGGAGATC 425
|||||

RESULT 9
AF318360 2820 bp mRNA linear HTC 01-JAN-2002
LOCUS AF318360 Homo sapiens pp7518 mRNA, complete cds.
ACCESSION AF318360
VERSION AF318360.1 GI:18027811
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2820)
Mammalia; Eutheria; Primates; Catarrhini: Hominoidea; Homo.
AUTHORS Zhou,X.M., Zhang,P.P., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T.,
Wan,D.F. and Gu,J.R.

TITLE Novel human cDNA clones with function of inhibiting cancer cell
growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2820)
Zhou,X.M., Zhang,P.P., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T.,
Wan,D.F. and Gu,J.R.

TITLE Direct Submission
JOURNAL Submitted (02-NOV-2000) National Laboratory For Oncogenes & Related
Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai
200032. P. R. China

FEATURES
source Location/Qualifiers
1..2820
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2820
/gene="pp7518"
1313..1621
/gene="pp7518"
/codon_start=1
/product="unknown"
/protein_id="AAL55867.1"
/db_xref="GI:18027812"
/translation="MVAGVGSGSWPHWVGLETLPLGLVKSKYKIAKILLVQSPITFL
ATGSIPLDHFSEMEKGGAGCGMEVHPLPSSGISQTLLESTONLHSCADITGCLLL
"

BASE COUNT 734 a 668 c 850 g 568 t
ORIGIN
Query Match 7.0%; Score 75; DB 11; Length 2820;
Best Local Similarity 100.0%; Pred. No. 6,5e-27;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 tcccagctactcggaggctgaacgaagaatcgccttgaacccggagcgaacgaatc 848
|||||
Db 955 TCCAGCTACTCGGAGGCTGAGCGAGGAGAAATGCTTGAACCGGCGGAGGTTG 1014
|||||

QY 849 atgaacgaatc 863
|||||

Db 1915 AGTGAGCGGAGATCA 1029
|||||

RESULT 10
BE061830 242 bp mRNA linear EST 05-JUN-2000
LOCUS BE061830
DEFINITION BC1-BF0254-271159-014-g11 BF0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE061830
VERSION BE061830.1 GI:8406480
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 242)
Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini: Hominoidea; Homo.

AUTHORS Dias Neto,E., Garcia Correa,P., Veronesi,Aurelio,S., Briones,M.R.,
Nadal,M.A., da Silva,W.F., Zebrowski,A., Bordin,S., Costa,F.P.,
Goldman,G.H., Carvalho,A.P., Marukawa,A., Bahia,G.S., Simpson,A.H.,
Brunstein,A., de Oliveira,P.S., Bortolotto,J., Jannone,C.V., Pabre
M.J., Soares,F., Brentani,R.B., Reiss,E., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7) 4191-4196. (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 05509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704522
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/ICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/getfasta.pl?l=512-BC1-BF0254-271
199-014-g11&t=3-199-11-27&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 76.
Location/Qualifiers

1..242
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BF0254"
/dev_stage="Adult"
/note="igrid: Breast; Vector: puc18; Size:1; Small: 512; 2;
Small: A mini-library was made by random products derived
from ORFESTES ppc (U.S. Letters Patent application No. 196
0716 - Ludwig Institute for Cancer Research) profiles
into the ppc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 68 a 54 c 73 g 47 t
ORIGIN
Query Match 6.9%; Score 74; DB 9; Length 242;
Best Local Similarity 100.0%; Pred. No. 3,2e-26;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 tcccagctactcggaggctgaacgaagaatcgccttgaacccggagcgaacgaatc 848
|||||
Db 106 TCCAGCTACTCGGAGGCTGAGCGAGGAGAAATGCTTGAACCGGCGGAGGTTG 145
|||||

QY 849 atgaacgaatc 862
|||||

Db 166 AGTGAGCGGAGATC 179
|||||

RESULT 11
A1053827/c 359 bp mRNA linear EST 27-JUL-1998
LOCUS A1053827/c
DEFINITION q170607.x1 NCICGAP Gr26 Homo sapiens cDNA clone IMAGE146143
similar to contains A19 repetitive element; contains element L145 L1
repetitive element ; mRNA sequence.

ACCESSION A1053827
VERSION A1053827.1 GI:3321614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 358)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 444 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham.
Location/Qualifiers
1. 358
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1861813"
/clone_lib="NCI-CGAP_Ov26"
/sex="female"
/tissue_type="papillary serous carcinoma"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pAMP1; mRNA made from
papillary serous ovarian carcinoma, cDNA made by oligo-dT
priming. Directionally cloned. Size-selected on agarose
gel, average insert size 500 bp. Primary library,
non-amplified."

BASE COUNT 66 a 110 c 87 g 95 t
ORIGIN

Query Match 6.9%; Score 74; DB 9; Length 358;
Best Local Similarity 100.0%; Pred. No. 3e-26;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 813 caggagaatcgcttgaacccggaggcggaggttgcagtgaggcggagatcacaccatgc 872
|||||
Db 105 CAGGAGAAATCGCTTGAACCCGGGAGGGAGGTTGCAGTGAGCCGAGATCACACCCTGC 46
|||||

QY 873 actccagcctgggc 886
|||||

Db 45 ACTCCAGCCTGGGC 32

RESULT 12
A0818830
LOCUS A0818830 441 bp DNA linear GSS 26-AUG-1999
DEFINITION HS_5014_B2_H06_SP5E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=590 Col=12 Row=P, DNA sequence.
ACCESSION A0818830
VERSION A0818830.1 GI:5781223
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 441)
AUTHORS Mahairas, G.C., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and
Hood, L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and

JOURNAL scanning the human genome
MEDLINE Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
COMMENT 99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (<http://info@resgen.com>). BAC end Web Server:
<http://www.htsc.washington.edu>
Plate: 590 row: P column: 12
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 441.
Location/Qualifiers
1. 441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=590 Col=12 Row=P"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site: 1; EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
135 a 99 c 111 g 93 t 3 others
BASE COUNT
ORIGIN

Query Match 6.9%; Score 74; DB 12; Length 441;
Best Local Similarity 100.0%; Pred. No. 2.9e-26;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 tccagctactcggaggcgtgaggcaggaggaatcgcttaacccggaggcggaggttc 848
|||||
Db 193 TCCAGCTACTCGGAGGCTGAGGAGGAGATCGCTTGAACCCGGGAGCGGAGTTGC 252
|||||

QY 849 agtgagccgagatc 862
|||||

Db 253 ACTGAGCCGAGATC 266

RESULT 13
A072006/c
LOCUS A072006 489 bp mRNA linear EST 08-MAR-2000
DEFINITION WS57907.x1 NCI-CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2501340.3
similar to contains Alu repetitive element; contains element LTR8
repetitive element ;, mRNA sequence.
ACCESSION A072006
VERSION A072006.1 GI:6027004
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 489)
AUTHORS NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute/ National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

CA). Note: this is a NIH_MGC Library."

BASE COUNT	126 a	156 c	163 g	149 t
ORIGIN				
Query Match	6.8%; Score 73; DB 10; Length 594;			
Best Local Similarity	100.0%; Pred. No. 8.8e-26;			
Matches	73; Conservative	0; Mismatches	0; Indels	0; Gaps
	0;			
Qy	789	tcccagctactcggaggctgagcagcagagaatcgcttgaacccgggagcggaggttgc	848	
Db	157	TCCCAGCTACTCGGAGGCTGAGCAGGAGGAATCGCTTGAACCCGGGAGCGGAGGTTGC	216	
Qy	849	agtgagccgagat	861	
Db	217	AGTGAGCCGAGAT	229	

Search completed: July 5, 2002, 05:32:58
Job time: 2926 sec

Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 293265
Center clone name: CIT-HSPC_281H14

Summary Statistics
Consensus quality: 123959 bases at least Q40
Consensus quality: 133398 bases at least Q30
Consensus quality: 134736 bases at least Q20
Estimated insert size: 135340; agarose-fp estimation
Estimated insert size: 136595; sum-of-ontigs estimation
Quality coverage: 4.95 in Q20 bases; agarose-fp estimation
Quality coverage: 4.91 in Q20 bases; sum-of-ontigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1
2163: contig of 2162 bp in length
2263: gap of unknown length
2263: contig of 2961 bp in length
5224: gap of unknown length
5324: contig of 9876 bp in length
15199: gap of unknown length
15200: contig of 7013 bp in length
22312: gap of unknown length
22413: contig of 14695 bp in length
37107: gap of unknown length
37207: gap of unknown length
43863: contig of 6656 bp in length
43963: gap of unknown length
46963: contig of 3002 bp in length
47066: gap of unknown length
49477: contig of 2412 bp in length
49577: gap of unknown length
54523: contig of 4946 bp in length
54623: gap of unknown length
68414: contig of 13791 bp in length
68514: gap of unknown length
72850: contig of 4336 bp in length
72950: gap of unknown length
88213: contig of 15263 bp in length
88313: gap of unknown length
98538: contig of 10225 bp in length
98638: gap of unknown length
107774: contig of 9136 bp in length
107874: gap of unknown length
115145: contig of 7271 bp in length
115245: gap of unknown length
118552: contig of 3307 bp in length
118553: gap of unknown length
128198: contig of 9546 bp in length
128298: gap of unknown length
130123: contig of 1825 bp in length
130124: gap of unknown length
130224: contig of 3815 bp in length
134039: gap of unknown length
134138: gap of unknown length
136470: contig of 2332 bp in length
136471: gap of unknown length
136571: contig of 975 bp in length.
Location/Qualifiers
1..137545

FEATURES
source
1..137545
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CITC-281H14"
/clone_lib="CalTech human BAC library C"

BASE COUNT 37277 a 32308 c 31466 g 34494 t 2000 others
ORIGIN
Query Match 61.2%; Score 4797; DB 2; Length 137545;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 7717; Conservative 0; Mismatches 14; Indels 20; Gaps 17;
QY 1 ctcgagcccgaggttcaagaccagcctgggaaacatagggagacccctctctccaca 60
Db 24854 CTCGAGCCCGAGGTTCAAGACGAGCCTGGGAAACATAGGAGACCCCTCTCTCTCCACA 24913
QY 61 aaaaatttaaaactagccaggtgtgtggcaaacctgtatgccagctactcagaag 120
Db 24914 AAAAAATTTAAAACTAGCCAGGTGTGTGGCAAAACACTGTAGTCCAGCTACTCAGAAG 24973
QY 121 gctgaggtggaggatcacttgagcctgaaagttagagctacagtgagcgtgatcac 180
Db 24974 GCTGAGGTGGAGGATCACTTGAGCCTGGAAAGTAGAGGTACGTGAGCCGCTGATCACA 25033
QY 181 ccactgcactccagcctgggagacagagtggagaccctgtcaataataaacacaaat 240
Db 25034 CCACCTGCACCTCCAGCCTGGGAGACAGAGTGAGACCCCTGTCAAATAAATAACAACAAT 25093
QY 241 aatgattaaaaactaaaaactaaattttatgtatatttcaactgttatttttaaagatt 300
Db 25094 AATGATTAAAAATACTAAAACTAAATTTATGCTATTTTTCACCTTGATTTTGTAAAGATT 25153
QY 301 tttaaatgaaattcccaattgcttccagaagattgttcaaatatataccacatt 360
Db 25154 TTTAAATGAAATTTCCCAATTTGCTTCCAGAAGATTGTTCAAATATATACCCACATT 25213
QY 361 tcactcatgttctcttcctgaacagcagcaatcagaaaaactccctgggaagaggcagg 420
Db 25214 TCACCTCATGTTCTCTCTCTGAAACAGCAGCAATCAGGAAAACTCCCTGGAAGAGCGAGG 25273
QY 421 cttagactgaagattttaaaaggggtaggcctcagctcctccagggtttacactgtgc 480
Db 25274 CTTAGACTGAGATTTTAAAGGGGGTAGGCCTCAGCTCTCTCTTCCAGGTTTACACTGTGC 25333
QY 481 atgtttccaaactcaagaatttacactctctggtgtgcatgtctgtaagatctgac 540
Db 25334 ATGTTTCCAACTCAAGAAATTTACACTCTTCTGTTGGATTGCTCTGTAAAGATCTGAC 25393
QY 541 ccactactatgtattaaaaagggatgcataataagaattcagccctctctcttaaaatcc 600
Db 25394 CCACCTACTATGTATTAAAAAGGGATGCATGATAATGAATTCAGCCCTCTCTGTAAAAATCC 25453
QY 601 aaagggctctattgcagtttccccatttaattgggtgcataaaatattcttgggaaggac 660
Db 25454 AAAGGGTCTTATTGCGATTTCCTCCCATTTAAATGGGTCATTAAATATTCTTGGGAAGGAC 25513
QY 661 aaagctttagtttaactatgagaaaaaagcagagacccagccctggattctgtctcaag 720
Db 25514 AAAGCTTTAGTTAACTATGAGAAAAACAAGCAGAACCCCTGGATTCTGTCTTCAAAG 25573
QY 721 attttaccatgttggcagcctggtagtccagagcccaagaaaaatccccagccacagat 780
Db 25574 ATTTTACCATGTGGCAGCCCTGGTAGTCCAGAGCCCAAGAAAAATATCCAGCCACAGAT 25633
QY 781 accctagatgtagactagcagtgctacaacctcaaggtcagaagtgtgctactagaccag 840
Db 25634 ACCCTAGATGTAGACTAGCAGTGTCTACAACTCAAGGTCAGAAGTATGTCACCTAGACCAG 25693
QY 841 agccaaaaataggtgctatlatcatlaagagagttaaaatgcaaacacacagacaggtgac 900
Db 25694 AGCCAAAAATAGTGCTATATCATTTAAGAGAGTAAAAATGCAAAACACACAGACGGTGAC 25753
QY 901 attattcaataaagcataataacccacaggggactctctatctgaatatatcaagaactct 960
Db 25754 ATTATTCAATAAAGCATAATAACCCACAGGGGACTCCTATCTCTCAATATATCAAGAAGACTCT 25813
QY 961 cactaatcaataagaaaaaggcaaaagatttaacaggcacttcacaaaaaaagtattat 1020

Db 2821 ATTTTCCCGCAGCGGACAAAGAGGATTTCGCTCAACAAGTAGCTGTAAACAAGATTC 2880
Qy cctattttcgtgtagaataaaagagcctccgaggagacatcgccc 3517
Db CCTATTTTCGTGTAGCAAAAAAAGAGAGCCCTCCGGGAGAGACATCGGCC 2940
Qy taatatttccagatggcggttcaagcggttgagagtttgtctctaccagc 3577
Db TAATATTTCTCCAGATGGCGGCTTC-AGCGCGTTTGAGAGTTTGCTCTCTACCAGC 2999
Qy ctcgggttcaggccccccacccctcctcctgctcccgcccttctctccacccctcc 3637
Db CTCGGGTTCTAGGCCCCCGCACCCCTCATCTGGCTCCGCCCCCTCTCTCCACCCCTCC 3059
Qy ggaccctaaaggcggggcccgaagcgagggcgctgcgcctgaccccgagcgaa 3697
Db GGACCCCTAAGGGGCGGGGCCCAAGCGGAGGGCGCTGGCCTGACCCCGAGCGGAA 3119
Qy gggcccagctagtcctaatgcgggtggcgctcctttgacaggcgcggttgggac 3757
Db GGGCCCCAGTCTAGTCTCTAATGCGGGTGGCGTCTCTTTGACAGCGCGGCTTTGGGAC 3179
Qy aacagcgggagagagataaggtagacataccagagagatttgtgcgcgtgatac 3817
Db AACAGCGGGGACGAGATAAGGTGACATACAGAGCAGATTTGGTGGCGCGCTGATAC 3239
Qy tctctcccgacagaaacgcgagctatttaaagaccctcctcctgattacttcttc 3877
Db TCCTCTCCCGACAGAAACGCGGAGCTATTTAAAGACCCCTATCGATTACTTATCTTC 3299
Qy ct---ggaaagcttcgagagacaaagatgttccctgccttaagacacaaaggcca 3934
Db CTGGAGGAAAGCTCTTCGCGAGAGACAAAAGATGTTCCCTGCCTTAAGACACAAGGCCA 3359
Qy caaacggaggtctgcacagcgacgcacaaattcggcgcggggaaagcaaaacacact 3994
Db CACAACGGAGGGTCTGCACAGGCGACGACAAATTCGGCGCGGGGAAAGCAAAACACACT 3419
Qy gacgcttagagtgcacaaacgtgtgttccagagagctccagagtgcgagggagcg 4054
Db GACGCTTAGAGTGACAAACGTGTGTTCACAGAGCAGCTCCAGAGTGGCGGAGGACG 3479
Qy ctggggcgcgaggggccaccagtagtgctctctgtgccccttggaagtcttttttc 4114
Db CTGGGCGCGCGAGGGGACCCACAGTATGTCTCTGTGCGCTTGGAAAGTCTTTTTC 3539
Qy accgtatgcgttaaaacacgcacacagagaaagtgaactgtcacttagggcgctgt 4174
Db ACCGTATGCGCGTAAACACGACACAGAGAAAGTGACTGTGCACCTTAGGGCGCCTGT 3599
Qy gtgtaccggtgtcttttagcgaatttaaagcacatcagggcgcgccatggctcacgc 4234
Db GTGTACCGGTGCTTTTAGCGAATTTAAACACATCAGGCGCGGCGCATGGCTCAGCG 3659
Qy ctgtaatccagcacttagaggccagggcgggccgatacctgaggtcggagttcga 4294
Db CTGTAATCCCAAGCACTTTAGAGGCGCAGGCGGCGCGCATCACTGAGTGGGAGTTCGA 3719
Qy caccagctggcgaacatggtgaacccctgtctcaaaaaatacaaaaaattagcggg 4354
Db CACCAGCTTGCCCAACATGGTGAACCCCTGTCTCTACAAAAAATACAAAAATTAGCCGGG 3779
Qy catggtgatgcgtgctgtgatccagctactcggaggtcgagcgaggaatacgttg 4414
Db CATGGTATGCTGCTGTGATCCAGCTACTCGGGAGGCTGAGGACGAGAGATCGCTTG 3839
Qy aacccggggagcgaggttgagtgagccgagatacaacacactgcactccagcctggcg 4474
Db AACCCGGGAGCGGAGGTGTCAGTGAGCCGAGATCACACCACTGCACCTCCAGCCTGGCGC 3899
Qy acaagagcgaattccgctct-aaaaataaaaaataaaataaaatga--taattaaagcca 4531
Db ACAAGCGGAATTTCCGCTCTCAAAAAATAAATAAATAAATAAATAAATAAAGACACA 3959

Qy tcaact-cacattcaaaagc-ggttactggtggttga-atgtat-ccatagacacaggtc 4532
Db TCAACTCCACATTCAAAGCGGTTACTGTGGTGTAGATGTATCCATAGACACAGGTC 4019
Qy taaaatgtaaacgctccattgtgctccttttaaggcgctgaaatgtctgaaactgtcatgt 4647
Db TAAAATGTAAAGCTCCATTGTGCTCTTTAAGGGCTTGAATGTCTGAACGTCTATGT 4079
Qy gtacacttaaaatggaatgtcaaacagaccctttctagcgctcgtttcgtctct 4707
Db GTACACTTAAAGATAGGATGTCAACAGACCCCTTTCTAGCGCTCGTTCGTGTCT 4139
Qy gaatcccgcatcttcgccaattgttgagcgagaaacgcccctcgcaaaagcgagctg 4767
Db GAATCCCCGCAATTCGCCAATTTGCTTGAGCGCAGAACGCCCTCCCGAAGAGCGGCTG 4199
Qy ctgatacccgacttctcgcgtatcgcgagcttgttgccctcggggtcccccgctgccat 4827
Db CTGATCCCGACTTGTCTCGGTATCGCGAGCTTGTGTGGCTTCGCGTCCCGCTGCCAT 4259
Qy gccccgggagctctccacagacacccgcttgcgcgcaattatacagagactgaatgggtt 4887
Db GCCCGCGGAGGCTCTCCACAGACACCGCTTGCGCGCAATTTATACGAGACTGAATGGGTT 4319
Qy ttttgggtgtgtgtgcaaacacaacaaattgtcagctcgtgttcaaatgqctccgc 4947
Db TTTTGGTGTGTGTGCAACACACAATTTGTACGTGCTGTTCACATGCGCTCCGCC 4379
Qy gggcggtgaaacttggctgcgtgtaacgcacagcaggttggaggggcacaccccgagga 5007
Db GGGCGGTGAAACTTGGCTGCGGTAAACGCACAGCAGGTTGGAGGGGCACACCCGGAAGGA 4439
Qy aggaagagcgagggaaagcgcgacccctagggccgctggccagcgtttccagca 5067
Db AGGAAGAGCGAGGAGGAAAGGCGCGACCTAGGCGCTGCGCGAGCGCTTTCAGCA 4499
Qy tcaattcagcactgagcgcgcgacagcagggctggggctcccggaaagtctcgcc 5127
Db TCAATTCAGCACTAGCGCGCCGACGACACAGAGGCTGGGGCTCCCGAGGTGCGGCC 4559
Qy agccggggttggcgagcgcgagagctgcccggtggtaggtcgactcttccactc 5187
Db AGCCCGGGTTTCGGCGAGAGCGCGAGGCTGCCGGTGGTAGGTGCGACTCTTTCACCTC 4619
Qy tccggggagcgcgccgacaccccaaccccaagcgtgcgctcgcccgccgctgg 5247
Db TCCGGGAGCGCGCGCGGACGACCCCAACCCACCCGC-AGCGCTGCGTCCGCCCGGCTGG 4678
Qy tcccccgcggggcacaaaacagcgcgagttcgccagctctcttccccaaacctgaa 5307
Db TCCCCCGCGGGGCACAAAAACAGCGCGAGTTGCCAGCTCTCTTTTCCCAACCTGAA 4738
Qy ccgcaagccgaaggttcttccaaagtcggttccccgggttcaaccccccgggcag 5367
Db CGGCAAGCCGAAGGTTCTTCCAAAGTCGCGGTTCCCGGGGTTTACACCCGCGGGCAG 4798
Qy gcgcaaacagccccagacacaccatttctcctcactgtatctgtgtgtccatc 5427
Db GCGCGAACCCAGCCCCAGGACAACCAATTTTCTTCTTCACTGTATCTGAGTGTGTGCCATC 4858
Qy tgaactcgaatgtcactgatcttcccgactgtgacctccagcgaggggactccgagaaac 5487
Db TGACTCGAATGTACCTGATTTTCCAGCTGTGACCTCCAGCGACGGGACTCCGAGGAAC 4918
Qy tgattccagcgtctcgattctctcgcctctccgcccgttttggctgaaagcggtttgcag 5547
Db TGATTCCAGCGCTCGATTCTCTCCGCCCTCTCCGCCGCTTTTGGCTGAAGCGGTTTGCAG 4978
Qy ccgctcgggcgaagaggggtgggagtggtgcagccacagccccagccccagagagaaga 5607
Db CCGTCCGGGCAGAAAGGGGTGGGATGTGGCAGCCACCGCCCCCAGCCAGAGAGAAAGA 5038

QY 5608 gdnqaaatfaacqcdaaagacacccgaadttctdaaagcactccctcgaatcctcga 5667
DB 5609 ggaacaaatfaacggcaaaaggacacccgaagttctaaaggcactccctcgaatcctcga 5668
QY 5668 atccqaaacaaacctaacartaatcttcaaaagcgaatcatatcactaatccagacaaa 5727
DB 5699 atccgagcgaacccctaacactagtttgaaggcggaatcatatccactaatccagacaaa 5158
QY 5728 ttcgggttggaacacatactcccagagccctaaqaaactacttacaacaaacaaac 5787
DB 5159 ttgcgggttggaacacatactcccagagccctaaqaaactacttacaacaaacaaac 5218
QY 5788 tgaacaaacaaatcgaagaggtttgaacactaatctcactcgaacaaatataata 5847
DB 5219 tgacaaacgaacaaatgcgaagaggtttggaacgttaattctctcgaacaaatataata 5278
QY 5848 tatatacatcctcataatattgttttaatttgcgaacaaacaaactctcgaagataata 5907
DB 5279 tataatacatcctcataatattgttttaatttgcgaacaaacaaactctcgaagataata 5338
QY 5908 ttttttaaacccagttgcagcttggaagagagtgagattagctgaagagggagagagaa 5967
DB 5439 ttttttaaacccagttgcagcacttggagagagtgagattagctgaagagggagagaa 5398
QY 5968 gcaatttgaggtgacgtaaatgtttttgtatcttgatattagtgagctgtatagaggtg 6027
DB 5499 gcaatttttgaggtgacgtaaatgtttttgtatcttgatattagtgagctgtatagaggtg 5458
QY 6028 cactacaagttgcaagactcatcgaactgtacacttttggcttagtactataactc 6087
DB 5459 cacatctcaagttgcaagactcatcgaactgtacacttttggcttagtactataactc 5518
QY 6088 aataaadtgaattttaaacctaaataagccagatcaacaggtttccctggagtgagctgag 6147
DB 5619 aataaadtgaattttaaacctaaataagccagatcaacaggtttccctggagtgagctgag 5578
QY 6148 aagagcttggaacactttacattgaatcctcctcttaggcagttcggtttggtttggtt 6207
DB 5579 agagggcttggaacactttacattgaatcctcctcttaggcagttcggtttggtttggtt 5638
QY 6208 tatctcagatattattttatccaaaatatatactaatgaacagatgaactaat 6267
DB 5639 tggtcttatagatattattttatccaaaatatatactaatgaacagatgaactaat 5698
QY 6268 gtaaacacttgtaagaaacacaaagacgaaagacagacactgagcactcctat 6327
DB 5699 gtanaacacttgtaagaaacacacaaagacgaaagacacactgagcactcctat 5758
QY 6328 tagaagataaagataagcactcacttgcgaagcicalaaatatatttggagataaatat 6387
DB 5759 tagagggataaagataaagcactcacttgcgaagcicalaaatatatttggagataaatat 5818
QY 6388 cgttcgcacttggtttttgtgtttttgtgtcctatgtgttaacatcaacgaaataaag 6447
DB 5819 cgttcgcacttggtttttgtgtttttgtgtcctatgtgttaacatcaacgaaataaag 5878
QY 6448 acccaaacctatccagtgatcagttcagatcagtgagtgagtgagtgagtgagtgagtgag 6507
DB 5879 acccaaacctatccagtgatcagttcagatcagtgagtgagtgagtgagtgagtgagtgag 5938
QY 6508 ctactgaagccacagcgtgggagagcacaatcctcctcagatgttgaagtcacagcag 6567
DB 5939 ctactgaagccacagcgtgggagagcacaatcctcctcagatgttgaagtcacagcag 5998
QY 6568 taagaacacacagtcctccttgcagcaagatgtgtgagcagcagcagcagcagcagcag 6627
DB 5999 taagcaacacacagtcctccttgcagcaagatgtgtgagcagcagcagcagcagcagcag 6058
QY 6628 ggcagagcccgagagagcctcagatctnaccgagcagatgaagagtggtgccctcagag 6687
DB 6059 ggcagagcccgagagagcctcagatctnaccgagcagatgaagagtggtgccctcagag 6118
QY 6688 ctgagagcctcgtgagcctgaga-gctcga-gctcga-gctcga-gctcga-gctcga-gctcga 6745

DB 6119 ctgagagcctcgtgagcctgagagaggtctgaggtgtgtgcaatgtgtgagagagagagagag 6178
QY 6746 ttcgaag 6805
DB 6179 ttgag 6248
QY 6806 tatagaatatttccatttaccacctacacccagacccagacccagacccagacccagacccag 6865
DB 6239 tatgaatatatttccatttaccacctacacccagacccagacccagacccagacccagaccc 6298
QY 6866 ttttgaag 6925
DB 6299 ttttgaag 6358
QY 6926 gctgaactctcgaag 6985
DB 6359 gctgaactctcgaag 6418
QY 6986 actgaactaaactcctcgtcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 7045
DB 6419 actgaactaaactcctcgtcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 6478
QY 7046 gctgaactctcgaag 7105
DB 6479 gctgaactctcgaag 6548
QY 7106 gctgaactctcgaag 7165
DB 6549 gctgaactctcgaag 6598
QY 7166 gctgaactctcgaag 7225
DB 6599 gctgaactctcgaag 6658
QY 7226 gctgaactctcgaag 7285
DB 6659 gctgaactctcgaag 6718
QY 7286 gctgaactctcgaag 7345
DB 6719 gctgaactctcgaag 6778
QY 7346 gctgaactctcgaag 7405
DB 6779 gctgaactctcgaag 6838
QY 7406 gctgaactctcgaag 7465
DB 6839 gctgaactctcgaag 6898
QY 7466 gctgaactctcgaag 7525
DB 6899 gctgaactctcgaag 6958
QY 7526 gctgaactctcgaag 7585
DB 6959 gctgaactctcgaag 7618
QY 7586 gctgaactctcgaag 7645
DB 7019 gctgaactctcgaag 7078
QY 7646 gctgaactctcgaag 7705
DB 7079 gctgaactctcgaag 7138
QY 7706 gctgaactctcgaag 7735
DB 7139 gctgaactctcgaag 7168

||||| 55236 ATGGAAGCAGGAGGAGACTGAGAGTCTGTTATTGTTATTGTTATTATTTATTTT 55177
Qy 2999 taaaaactgat-cagccgactttgaatacagaaaaatgaaagatgagagatttgata 3057
|||||
Db 55176 TAAAAAATGGATACAGCCGACTTGAATACAGAAAATGAAAAATGAGAGATTGCGATA 55117
Qy 3058 acagcgttgacgtctgaaggggcccagggccctagcgctgggtggggcacctagaaaca 3117
|||||
Db 55116 ACAGCGTTGGACGCTGAAGGGGCCACAGGGCCTAGCGGCTGGTGGGGCACCTAGAAACA 55057
Qy 3118 cttctgctgcagatcgcggaagggttagcaacagaaaggggtgccttaggctgcccacag 3177
|||||
Db 55056 CTTTGGCTGCAGATCGCGGAGGTTAGCCACAGAAAGGGTGCCTTAGGCTGGCCACAG 54997
Qy 3178 ggcctttgctgactgaaggaacagccttgccggaaccttcttcccctctgccttga 3237
|||||
Db 54996 GGCCTTTGCTGACTGAAGACACAGCCTTGGCGGCACCTTCTTCCCTCTGCCCTGCA 54937
Qy 3238 ctcgggccccgcggagtcagagctgacttgctgcaggttggggagagacagagctag 3297
|||||
Db 54936 CTCGGGCCCCCGAGTTCAGAGCTGACTTGTGTCAGGTTGGGGAGAGACAGAGGCTAG 54877
Qy 3298 gacgttgcaaaacctcacctcgtcgcagtcgcgaagtaaacitggaccgcgagcac 3357
|||||
Db 54876 GACGTTGCGGAACCTCACCTGCTCGCAGTCCGGAAGTAACCTTGGACCCGCGCAGGCAC 54817
Qy 3358 ttctctaaagtccaagctgcctctctgaagataaacctgatttctcccggaacgcgac 3417
|||||
Db 54816 TTCTTAAGTCCAAGCTGCCCTCTCTGAAGAATAAACCCTGATTTCCTCCGAGCGGCAC 54757
Qy 3418 aaaggaggttcgtctcaaacactagctcgtgaacaaagattccctatttctggttagaa 3477
|||||
Db 54756 AAAGGAGGATTTCGCTCACTCACTAGCTGTTAAACAAAGATTCCCTATTTCGTGGTTAGAA 54697
Qy 3478 aaaaaaaagaaagccctccgggagagacatgcacctaatatttctccagatgg 3537
|||||
Db 54596 AAAAAAANAAGAACCCCTCCGGGAGAGACATGCGCCCTAAATATTTCCTCCAGATGGG 54637
Qy 3538 ccgggttcaagcggcttgagagttgctctctaccagcctcgggttctaggccccccg 3597
|||||
Db 54636 CCGGTTTC-AGCGGTTTGAGAGTTTGCTCTCTACAGGCTCGGTTAGGCCCCCGC 54578
Qy 3598 cacctcatcctgctccggccctctctccacctcccgagaccctaaaggggcgcg 3657
|||||
Db 54577 CACCTCATCTGCTCCGCCCTCTCTCCACCTCCCGACCCCTTAAAGGGCGGCG 54518
Qy 3658 gggcccaagcggcgctgcctgaccccgagcgaaggcccaagtcctagttccta 3717
|||||
Db 54517 GGGCCCAAGCGGAGGCTCGGCTGACCCCGAGGAGGGCCCGCTAGTCTCCTA 54458
Qy 3718 atcgvggtggcgtctctttgacaggcggtttggggacacagcgggagcagagata 3777
Db 54457 ATCGGGTGGCGTCTCTTTGACAGCGCGCTTTGGGGACAACAGCGGGGAGAGATA 54398
Qy 3778 agtgacataccagagagattggtgcgcgctgatactcctctcccgagaaacg 3837
Db 54397 AGTGTCATACACAGAGAGATTGGTGGCGCGCTGATCTCTCTCCCGACAGAAACG 54338
Qy 3838 cggagctattaaagaccctatcgattacttattcttct---ggaagcttctgcg 3894
|||||
Db 54337 CGGAGCTATTATAAAGACCCCTATCGATTACTTTATCTTCTTGAGGAAAGCTTCTTCG 54278
Qy 3895 gagagacaaaagatttccctgccttaagacacaaagggccacacagcggaggtctgcaca 3954
Db 54277 GAGAGACAAAGATGTTCCCTGCTTAAGACACAAGGCCACACAGGAGGTTCGACA 54218
Qy 3955 ggcgacgcacaaattcgcgcgcgggaaagcaaaaaacacactgacttagtgcacaac 4014
Db 54217 GGCACGCACAAATTCGGCGCGGGGAAAGCAAAACACACTGACGCTTAGAGTGCACAAAC 54158
Qy 4015 gtatgtattccagacagctccagagtcggaaggacgctggggcgaggggacac 4074
|||||

Db 54157 GTGTGTGTTCACAGAGCAGCTCCAGAGTGCAGGAGCGCTGGGGCGCGAGGGGCAC 54098
Qy 4075 ccacagtatggtcttctgtgcccttgaaagtttttttcacccgtatgcgcgtaaaaaac 4134
|||||
Db 54097 CCACAGTATGTTCTGTGCGCTTGAAAGTTTTCACCGTATGCGCGTAAAAACAC 54038
Qy 4135 gcacacacagaaagtgcactgtgcacttagggccctgtgtgtacctccctgtcggttttag 4194
Db 54037 GCACACACAGAAAGTGACTGTGCACTTAGGGCGCTGTGTGTACCCGTCGCTTTAG 53978
Qy 4195 cgaatttaaacacacatacagccggcgccatgctcacgcctgtaatccacacactttag 4254
Db 53977 CGAATTTAAAGCACATCAGCGGGCGCATGGCTCAGCCCTGTAATCCAGCAGCTTAG 53918
Qy 4255 gagccgagggcgccgatcacctgaggtcgggagttcgacaccagcctggccaacatgg 4314
Db 53917 GAGCCGAGGGCGGCGGATCACCTGAGGTCGGAGTTTCGACACCAGCCTGGCCAAACATGG 53858
Qy 4315 tgaacctgtctctacaaaaatacaaaaattaccgggcatggtgctgctgctgtg 4374
Db 53857 TGAACCCCTGTCTCTACAAAAAATACAAAAATTAGCCGGCATGGTGTGCTGCTGTG 53798
Qy 4375 atcccagctactcggaggtgagcagagagatcgcttgaacccgggagcgaggttg 4434
Db 53797 ATCCAGCTACTCGGAGGCTGAGCAGAGAGTAATCGCTTGAACCCCGGAGGGAGTTG 53738
Qy 4435 cagtgagcagatcacaccactgcactccagcctcggcgcaagaagcgaattccgctc 4494
Db 53737 CAGTGAGCGAGATCACACCACCTGCACCTCCAGCCTGGGCGACAGAGCGAAATTCGCTCT 53678
RESULT 4
ALI36219/c
LOCUS Human DNA sequence from clone RP11-149B7 on chromosome 13, complete
DEFINITION sequence.
ACCESSION ALI36219
VERSION ALI36219.17 GI:12044488
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 167830)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Wall, M.
JOURNAL Direct Submission
Submitted (05-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Jan 8, 2001 this sequence version replaced gi:11863351.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-149B7 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm

misc_feature /note="assembly_fragment"
11354..15971
misc_feature /note="assembly_fragment"
16072..21310
clone_end:T7
vector_side:left
misc_feature 21411..32494
/note="assembly_fragment"
32595..47895
misc_feature /note="assembly_fragment"
47996..64361
/note="assembly_fragment"
64462..81134
/note="assembly_fragment"
81235..105677
/note="assembly_fragment"
105778..135680
/note="assembly_fragment"
135781..174537
/note="assembly_fragment"
BASE COUNT 55112 a 33702 c 33796 g 50727 t 1200 others
ORIGIN

Query Match 1.3% Score 98; DB 2; Length 174537;
Best Local Similarity 100.0%; Pred. No. 1.5e-42;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4376 tccagctactcggagctgagcgagcaggaatcgcttgaaccggcgagctgttc 4435
|||||
Db 86180 TCCAGCTACTCGGAGCTGAGCGAGGAGAACTGCTTGACCGGGAGCGAGTTGC 86239
|||||
QY 4436 agtgagccgagatcacaccactgcactccagcctgggc 4473
|||||
Db 86240 AGTGAGCGAGATCACACCCTGACACTCCAGCCTGGGC 86277
|||||

RESULT 6
AP000871/c
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-709J7 map 11q23, WORKING
DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION AP000871.3 GI:11320823
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 189492)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 189,492 genomic DNA of 11q23
JOURNAL Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 189492)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (13-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Nov 22, 2000 this sequence version replaced gi:8119019.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: RP11-709J7

----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 184505 bases at least Q40
Consensus quality: 186253 bases at least Q30
Consensus quality: 187002 bases at least Q20
Insert size: 187692; sum-of-contigs
Quality coverage: 8.38x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
19 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 35490 contig of 35490 bp in length
35591 56409 contig of 20819 bp in length
56510 77235 contig of 20726 bp in length
77336 88610 contig of 11275 bp in length
88711 110084 contig of 21374 bp in length
110185 118234 contig of 8050 bp in length
118335 127661 contig of 9327 bp in length
127762 136754 contig of 9124 bp in length
136855 145978 contig of 100 bp in length
145979 146078 gap of 100 bp
146079 152076 contig of 5998 bp in length
152077 152176 gap of 100 bp
152177 159008 contig of 6832 bp in length
159009 159108 gap of 100 bp
159109 165698 contig of 6590 bp in length
165699 171331 contig of 5533 bp in length
171332 171431 gap of 100 bp
171432 175338 contig of 3907 bp in length
175339 175438 gap of 100 bp
175439 179667 contig of 4229 bp in length
1 35490: contig of 35490 bp in length
35491 35590: gap of 100 bp
35591 56409: contig of 20819 bp in length
56410 56509: gap of 100 bp
56510 77235: contig of 20726 bp in length
77236 77335: gap of 100 bp
77336 88610: contig of 11275 bp in length
88611 88710: gap of 100 bp
88711 110084: contig of 21374 bp in length
110085 110184: gap of 100 bp
110185 118234: contig of 8050 bp in length
118235 118334: gap of 100 bp
118335 127661: contig of 9327 bp in length
127662 127761: gap of 100 bp
127762 136754: contig of 8993 bp in length
136755 136854: gap of 100 bp
136855 145978: contig of 9124 bp in length
145979 146078: gap of 100 bp
146079 152076: contig of 5998 bp in length
152077 152176: gap of 100 bp
152177 159008: contig of 6832 bp in length
159009 159108: gap of 100 bp
159109 165698: contig of 6590 bp in length
165699 171331: contig of 5533 bp in length
171332 171431: gap of 100 bp
171432 175338: contig of 3907 bp in length
175339 175438: gap of 100 bp
175439 179667: contig of 4229 bp in length
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 35490: contig of 35490 bp in length
35491 35590: gap of 100 bp
35591 56409: contig of 20819 bp in length
56410 56509: gap of 100 bp
56510 77235: contig of 20726 bp in length
77236 77335: gap of 100 bp
77336 88610: contig of 11275 bp in length
88611 88710: gap of 100 bp
88711 110084: contig of 21374 bp in length
110085 110184: gap of 100 bp
110185 118234: contig of 8050 bp in length
118235 118334: gap of 100 bp
118335 127661: contig of 9327 bp in length
127662 127761: gap of 100 bp
127762 136754: contig of 8993 bp in length
136755 136854: gap of 100 bp
136855 145978: contig of 9124 bp in length
145979 146078: gap of 100 bp
146079 152076: contig of 5998 bp in length
152077 152176: gap of 100 bp
152177 159008: contig of 6832 bp in length
159009 159108: gap of 100 bp
159109 165698: contig of 6590 bp in length
165699 171331: contig of 5533 bp in length
171332 171431: gap of 100 bp
171432 175338: contig of 3907 bp in length
175339 175438: gap of 100 bp
175439 179667: contig of 4229 bp in length

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0555F01
----- Summary Statistics -----
Sequencing vector: M13; 81%
Sequencing vector: plasmid; 19%
Chemistry: Dye-terminator ET; 81% of reads
Chemistry: Dye-terminator Big Dye; 19% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 166484 bases at least Q40
Consensus quality: 171727 bases at least Q30
Consensus quality: 174384 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 179147; sum-of-contigs
Quality coverage: 3.63 in Q20 bases; agarose-fp
Quality coverage: 3.93 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as:
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1
* 1833: contig of 1833 bp in length
* 1834
* 1933: gap of unknown length
* 1934
* 3870: contig of 1937 bp in length
* 3871
* 6886: contig of 2916 bp in length
* 6887
* 6887: gap of unknown length
* 10052: contig of 3066 bp in length
* 10053
* 10152: gap of unknown length
* 14183: contig of 4031 bp in length
* 14184
* 14283: gap of unknown length
* 17263: contig of 2980 bp in length
* 17264
* 17363: gap of unknown length
* 17364
* 23188: contig of 5825 bp in length
* 23189
* 23289: gap of unknown length
* 23289
* 28005: contig of 4717 bp in length
* 28006
* 32605: gap of unknown length
* 32606
* 32705: contig of 4500 bp in length
* 38209: gap of unknown length
* 38210
* 38309: gap of 5504 bp in length
* 38310
* 44789: contig of 6480 bp in length
* 44790
* 44889: gap of unknown length
* 51559: contig of 6670 bp in length
* 51560
* 51659: gap of unknown length
* 51660
* 59185: contig of 7526 bp in length
* 59186
* 59285: gap of unknown length
* 59286
* 67326: contig of 8041 bp in length
* 67327
* 67426: gap of unknown length
* 73591: contig of 6165 bp in length
* 73592
* 82463: gap of unknown length
* 82464
* 82563: gap of unknown length
* 90712: contig of 8149 bp in length
* 90713
* 90812: gap of unknown length
* 90813
* 110037: contig of 19225 bp in length
* 110038
* 110137: gap of unknown length
* 110138
* 137265: contig of 27128 bp in length
* 137266
* 137266: gap of unknown length
* 137366
* 181047: contig of 43682 bp in length.

FEATURES

source

1. .181047
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-555F1"
1. .1833

misc_feature

misc_feature
/note="assembly_name:Contig13"
1934. .3870
/note="assembly_name:Contig14"
3971. .6886
/note="assembly_name:Contig15
clone_end:17
vector_side:right"
6987. .10052
/note="assembly_name:Contig16"
10153. .14183
/note="assembly_name:Contig17"
14284. .17263
/note="assembly_name:Contig18"
17364. .23188
/note="assembly_name:Contig19"
23289. .28005
/note="assembly_name:Contig20"
28106. .32605
/note="assembly_name:Contig21"
32706. .38209
/note="assembly_name:Contig22"
38310. .44789
/note="assembly_name:Contig23"
44890. .51559
/note="assembly_name:Contig24"
51660. .59185
/note="assembly_name:Contig25"
59286. .67326
/note="assembly_name:Contig26"
67427. .73591
/note="assembly_name:Contig27"
73692. .82463
/note="assembly_name:Contig28"
82564. .90712
/note="assembly_name:Contig29"
90813. .110037
/note="assembly_name:Contig30"
110138. .137265
/note="assembly_name:Contig31"
137366. .181047
/note="assembly_name:Contig32"
BASE COUNT 40954 a 47507 c 48668 g 42006 t 1912 others
ORIGIN

Query Match 1.2% Score 95; DB 2; Length 181047;
Best Local Similarity 100.0%; Pred. No. 7e-41;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4383 tactcgaggagtgaggcaggagaaatcgctgaacccggaggcgaggttcagtgagc 4442
|||||
Db 154633 TACTCGGAGGCTGAGGCGAGGAGAAATCGCTTGAACCCGGAGCGGAGGTTGCAGTGAGC 154574

Qy 4443 cgagatacacactgcactccagctggcgagaca 4477
|||||
Db 154573 CGAGATCACACACTGCACCTCCAGCCTGGCGGACA 154539

RESULT 9

AC092602 AC092602 166634 bp DNA linear HTG 19-JUL-2001
LOCUS Homo sapiens chromosome 2 clone RP11-134G21, WORKING DRAFT
DEFINITION
SEQUENCE, 3 unordered pieces.
AC092602 AC060810
AC092602.1 GI:14916187
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166634)
Waterston,R.H.
The sequence of Homo sapiens clone

```

FEATURES             

|                                                                                                         |                     |
|---------------------------------------------------------------------------------------------------------|---------------------|
| Unpublished                                                                                             |                     |
| 2                                                                                                       | (bases 1 to 166634) |
| WATERMARK                                                                                               |                     |
| Submitter: NCBI                                                                                         |                     |
| Submission: 19-JUL-2011                                                                                 |                     |
| Genome Sequencing Center, Washington                                                                    |                     |
| University School of Medicine, 4444 Forest Park Parkway, St. Louis,                                     |                     |
| MO 63109, USA                                                                                           |                     |
| 4-Jul-19, 2001 this sequence version replaced gi:10194554.                                              |                     |
| COMMENT                                                                                                 |                     |
| -----                                                                                                   |                     |
| Genome Center                                                                                           |                     |
| Center: Washington University Genome Sequencing Center                                                  |                     |
| Center code: WUGSC                                                                                      |                     |
| Web site: <a href="http://genome.wustl.edu/asg/index.shtml">http://genome.wustl.edu/asg/index.shtml</a> |                     |
| -----                                                                                                   |                     |
| Project Information                                                                                     |                     |
| -----                                                                                                   |                     |
| Center project name: HNC014621                                                                          |                     |
| Drafting center: WMR                                                                                    |                     |
| -----                                                                                                   |                     |
| Summary Statistics                                                                                      |                     |
| -----                                                                                                   |                     |
| Sequencing vector: M13, 98%                                                                             |                     |
| Sequencing vector: plasmid: 62%                                                                         |                     |
| Chemistry: Iye-Primer ET: 9% of reads                                                                   |                     |
| Chemistry: Iye-terminator Bid Iye: 100% of reads                                                        |                     |
| Assembly program: Phrap: version 0.900419                                                               |                     |
| Consensus quality: 164307 bases at least Q40                                                            |                     |
| Consensus quality: 165043 bases at least Q60                                                            |                     |
| Consensus quality: 165484 bases at least Q80                                                            |                     |
| Insert size: 177600; annotated                                                                          |                     |
| Insert size: 167751; sum of contigs                                                                     |                     |
| Quality coverage: 9.50 in 920 bases; annotated                                                          |                     |
| Quality coverage: 19.07 in 920 bases; sum of contigs                                                    |                     |
| -----                                                                                                   |                     |
| * NOTE: this is a 'working draft' sequence. It currently                                                |                     |
| * consists of 3 contigs. The true order of the pieces                                                   |                     |
| * is not known and their order in this sequence record is                                               |                     |
| * arbitrary. Gaps between the contigs are represented as                                                |                     |
| * runs of N, but the exact sizes of the gaps are unknown.                                               |                     |
| * This record will be updated with the finished sequence,                                               |                     |
| * as soon as it is available and the accession number will                                              |                     |
| * be preserved.                                                                                         |                     |
| -----                                                                                                   |                     |
| 1 9488: contig of 9488 bp in length                                                                     |                     |
| * 9489 9588: gap of unknown length                                                                      |                     |
| * 9589 4431: contig of 24813 bp in length                                                               |                     |
| * 4402 4501: gap of unknown length                                                                      |                     |
| * 4502 166634: contig of 162133 bp in length.                                                           |                     |
| Location/Qualifiers                                                                                     |                     |
| 1..166634 "name:Sapiens"                                                                                |                     |
| /db_xref="taxon:9606"                                                                                   |                     |
| /chromosome="2"                                                                                         |                     |
| /clone="RP11-13421"                                                                                     |                     |
| 1..9488                                                                                                 |                     |
| /note="assembly_name:contig29"                                                                          |                     |
| 9589..4431                                                                                              |                     |
| /note="assembly_name:contig40"                                                                          |                     |
| clone_end:SP6                                                                                           |                     |
| vector_side:left                                                                                        |                     |
| 44502..166634                                                                                           |                     |
| /note="assembly_name:contig41"                                                                          |                     |
| clone_end:17                                                                                            |                     |
| vector_side:right                                                                                       |                     |
| BASE COUNT 44777 a 45842 c 38463 g 43387 t 260 others                                                   |                     |
| ORIGIN                                                                                                  |                     |
| -----                                                                                                   |                     |
| Query Match 1.1% Score 89; DB 2: Length 166634;                                                         |                     |
| Best Local Similarity 100.0%; Prod. No. 1,60-47;                                                        |                     |
| Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;                                             |                     |
| QY 4389 aaagatcgaacagacaaatcacttgaacccggaagccuagatcagccagat 4448                                        |                     |
| Db 9885g GGAGGGTGGAGCGAGCAAAATCGTTGAACCGCGGAGGCGAGGTGATGAGACGAGAA 98918                                 |                     |
| YY 4449 caaacactgaactcgaactctgaagcaga 4477                                                              |                     |


```


misc_feature /note="assembly_fragment"
39109..44224
/note="assembly_fragment"
44325..50232
/note="assembly_fragment"
50333..58422
/note="assembly_fragment"
58523..64440
/note="assembly_fragment"
64541..72727
/note="assembly_fragment"
72828..81027
/note="assembly_fragment"
81128..91377
/note="assembly_fragment"
91478..102460
/note="assembly_fragment"
102561..117037
/note="assembly_fragment"
117138..129181
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
129282..149602
/note="assembly_fragment"
149703..178508
/note="assembly_fragment"
BASE COUNT 45506 a 42854 c 42878 g 45067 t 2203 others
ORIGIN

Query Match 1.1%; Score 89; DB 2; Length 178508;
Best Local Similarity 100.0%; Pred. No. 1.6e-37;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4389 ggagcgtgagcagagagaatcgcttgaaaccggagggaggggttgagtgagcgcagat 4448
|||||
Db 96010 GGAGGCTGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 95951
|||||
Qy 4449 cacaccactgcactccagctggcgaca 4477
|||||
Db 95950 CACACCACCTGCACCTCCAGCCCTGGGGGACA 95922
|||||

RESULT 12
AC010984
LOCUS AC010984 191540 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-510C1 from 2, complete sequence.
ACCESSION AC010984
VERSION AC010984.7 GI:15668160
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 191540)
AUTHORS Du,F., Maupin,R. and Hawkins,M.
TITLE The sequence of Homo sapiens BAC clone RP11-510C1
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 191540)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 191540)
AUTHORS Waterston,R.H.
TITLE Direct Submission

JOURNAL Submitted (19-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 191540)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Sep 19, 2001 this sequence version replaced gi:14349340.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0510C01

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-397H17, 2000 bp overlap;
the clone sequenced to the right is AC023040. Actual start of this
clone is at base position 29090 of RP11-109E12.

FEATURES
Location/Qualifiers
1..191540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-510C1"
/clone_lib="RPCI-11"
55..455
/note="match to EST AA632464 (NID:92555878) np80e08.s1"
418..940
/note="similar to Homo sapiens EST AV732195
(NID:gl0849740)"
761..1010
/note="match to EST BF355538 (NID:gl1314612)"
795..1415
/note="match to EST BF107120 (NID:gl0889561)"
1028..1189
/note="match to EST BE350015 (NID:g9261868) ht07h01.x1"
1028..1385
/note="match to EST AL043002 (NID:g5422428)"
1034..1504
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:1188875.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7490

Center clone name: 209_E_9

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 146689 bases at least Q40

Consensus quality: 149124 bases at least Q30

Consensus quality: 149969 bases at least Q20

Insert size: 151000; agarose-fp

Quality coverage: 5.9 in Q20 bases; agarose-fp

Quality coverage: 5.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 . 9341: contig of 9341 bp in length
* 9342 9441: gap of 100 bp
* 9442 11373: contig of 1932 bp in length
* 11374 11473: gap of 100 bp
* 11474 47057: contig of 35584 bp in length
* 47058 47157: gap of 100 bp
* 47158 49227: contig of 2070 bp in length
* 49228 49327: gap of 100 bp
* 49328 54639: contig of 5312 bp in length
* 54640 54739: gap of 100 bp
* 54740 59242: contig of 4503 bp in length
* 59243 59342: gap of 100 bp
* 59343 65284: contig of 5942 bp in length
* 65285 65384: gap of 100 bp
* 65385 70923: contig of 5539 bp in length
* 70924 71023: gap of 100 bp
* 71024 78085: contig of 7062 bp in length
* 78086 78185: gap of 100 bp
* 78186 87419: contig of 9234 bp in length
* 87420 87519: gap of 100 bp
* 87520 97716: contig of 10197 bp in length
* 97717 97816: gap of 100 bp
* 97817 120620: contig of 22804 bp in length
* 120621 120720: gap of 100 bp
* 120721 150210: contig of 29490 bp in length
* 150211 150310: gap of 100 bp
* 150311 151865: contig of 1555 bp in length.
Location/Qualifiers
1 . 151865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"

FEATURES
source

/map="8"
/clone="RP11-209E9"
/clone_lib="RPCI-11 Human Male BAC"
1 . 9341
/note="assembly_fragment
clone_end:SP6
vector_side:left"
9442 . 11373
/note="assembly_fragment"
11474 . 47057
/note="assembly_fragment"
47158 . 49227
/note="assembly_fragment"
49328 . 54639
/note="assembly_fragment"
54740 . 59242
/note="assembly_fragment"
59343 . 65284
/note="assembly_fragment"
65385 . 70923
/note="assembly_fragment"
71024 . 78085
/note="assembly_fragment"
78186 . 87419
/note="assembly_fragment"
87520 . 97716
/note="assembly_fragment"
97817 . 120620
/note="assembly_fragment"
120721 . 150210
/note="assembly_fragment"
150311 . 151865
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 45019 a 30571 c 30486 g 44487 t 1302 others
ORIGIN
Query Match 1.1%; Score 87; DB 2; Length 151865;
Best Local Similarity 100.0%; Pred.No. 2.1e-36;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4387 cggaggctgagcgaggagaatcgcttgaaacccggaggcgaggttgagtcagtcgagc 4446
|||||
Db 107399 CGGAGGCTGAGCGAGGAGAAATCGCTTGAACCGGAGGCGGAGTTGCAGTGGCGGAG 107458
QY 4447 atcacaccactgcactccagctgggc 4473
|||||
Db 107459 ATCACACCACTGCACCTCCAGCCTGGGC 107485
RESULT 14
AL590113 AL590113 165292 bp DNA linear HTG 30-JAN-2002
LOCUS Homo sapiens chromosome 1 clone RP11-234D19, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL590113
VERSION AL590113.7 GI:18250776
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
AUTHORS Hall,R.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jan 19, 2002 this sequence version replaced gi:17154305.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute


```
QY 4449 c 4449
Db 309 c 309

RESULT 7
US-08-465-981-5/c
; Sequence 5, Application US/08465981
; Patent No. 5874415
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,981
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/324,465
; FILING DATE: October 17, 1994
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-465-981-5

Query Match 0.8%; Score 61; DB 2; Length 1725;
Best Local Similarity 100.0%; Pred. No. 6.3e-15;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4389 ggagcgtgagcagagaatcgcttgaaacccggaggcggaggttcagtgagccgaat 4448
|||||
Db 369 GGAGGCTGAGCAGAGAGATCGCTTGAACCCGGGAGGTTGCAGTGAGCCGAGAT 310
|||||

QY 4449 c 4449
Db 309 c 309

RESULT 8
PCT-US93-11915-5/c
; Sequence 5, Application PC/TUS9311915
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; APPLICANT: Abe, Miyako
; TITLE OF INVENTION: ENHANCER SEQUENCE FOR MODULATING
```

```
; TITLE OF INVENTION: EXPRESSION IN EPITHELIAL CELLS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11915
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/999,742
; FILING DATE: December 31, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/065W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1725
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US93-11915-5

Query Match 0.8%; Score 61; DB 5; Length 1725;
Best Local Similarity 100.0%; Pred. No. 6.3e-15;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4389 ggagcgtgagcagagaatcgcttgaaacccggaggcggaggttcagtgagccgaat 4448
|||||
Db 369 GGAGGCTGAGCAGAGAGATCGCTTGAACCCGGGAGGTTGCAGTGAGCCGAGAT 310
|||||

QY 4449 c 4449
Db 309 c 309

RESULT 9
US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
```



```
? SEQ ID NO 7
? LENGTH: 72604
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: modified base
? LOCATION: all n positions
? OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

Query Match          0.8%; Score 61; DB 4; Length 72604;
Best Local Similarity 100.0%; Pred. No. 4.7e-15;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4388 qqagagctgagcagagagaaatcgcttaaccggagagcagcagatgttcagatgagccgaga 4447
|||||
Db 15692 gggagctgagcagagagaaatcgcttaaccggagagcagcagatgttcagatgagccgaga 15951

QY 4448 t 4448
Db 15952 t 15952

RESULT 10
US-09-618-100B-4
? Sequence 3, Application US/08618100B
? Patent No. 6068976
? GENERAL INFORMATION:
? APPLICANT: Briggs, Michael R.
? APPLICANT: Alwerdt, Johan
? APPLICANT: de Vos, Piet
? APPLICANT: Staels, Bart
? APPLICANT: Creston, Glenn E.
? APPLICANT: Miller, Stephen G.
? TITLE OF INVENTION: MODULATORS OF ob GENE AND
? TITLE OF INVENTION: SCREENING METHODS THEREFOR
? NUMBER OF SEQUENCES: 48
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Lyon & Lyon
? STREET: 633 West Fifth Street
? CITY: Suite 4700
? CITY: Los Angeles
? STATE: California
? COUNTRY: U.S.A.
? ZIP: 90071-2066
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
? MEDIUM TYPE: Storage
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: IBM P.C. DOS 5.0
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/618,100B
? FILING DATE: March 19, 1996
? CLASSIFICATION: 514
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/558,588
? FILING DATE: October 30, 1995
? APPLICATION NUMBER: 08/510,534
? FILING DATE: August 2, 1995
? APPLICATION NUMBER: 08/418,036
? FILING DATE: April 5, 1995
? APPLICATION NUMBER: 08/408,534
? FILING DATE: March 20, 1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Warburg, Richard J.
? REGISTRATION NUMBER: 32,327
? REFERENCE/DOCKET NUMBER: 219/075
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (213) 489-1600
? TELEFAX: (213) 955-0440
? TELEX: 57-3510
```

```
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 10684 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? DESCRIPTION: Sequence between exon 1 and exon 2
? Patent No. 6068976
US-09-618-100B-3

Query Match          0.7%; Score 54; DB 3; Length 10684;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4389 aaagctgaaacagagagaaatcgcttaaccggagagcagcagatgttcagatgagccgaga 4446
|||||
Db 4657 GGAGGCTGAGGAGGAGAGAAATCGCTTGAACAGGAGAGAGAGATGTTCAGTGAAGAGAG 4714

RESULT 11
US-09-328-111-741/C
? Sequence 741, Application US/99328111;
? Patent No. 6262333
? GENERAL INFORMATION:
? APPLICANT: Endegre, Wilson G.
? APPLICANT: Steinmann, Kathleen E.
? APPLICANT: Astle, Jon H.
? APPLICANT: Burgess, Christopher C.
? APPLICANT: Businell, Steven E.
? APPLICANT: Carroll III, Eddie
? APPLICANT: Catino, Theodore J.
? APPLICANT: Dertli, Adnan
? APPLICANT: Ford, Donna M.
? APPLICANT: Lewis, Marcia E.
? APPLICANT: Monahan, John E.
? APPLICANT: Schlegel, Robert
? TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
? TITLE OF INVENTION: PRODOTS
? FILE REFERENCE: CDD-257 (US)
? CURRENT APPLICATION NUMBER: US/09/328,111
? EARLIER FILING DATE: 1999-06-08
? EARLIER APPLICATION NUMBER: US 60/084,801
? EARLIER FILING DATE: 1998-06-16
? NUMBER OF SEQ ID NOS: 850
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 741
? LENGTH: 579
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(579)
? OTHER INFORMATION: n = A,T,C or G
US-09-328-111-741

Query Match          0.7%; Score 54; DB 4; Length 579;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4382 ctactggagaaatgagcagcagagaaatcgcttaaccggagagcagcagatgttcagatgagccgaga 4436
|||||
Db 213 CTACTGGAGAGCTGAGGAGGAGAGAAATCGCTTGAATGAGAGAGATGTTCAGTGAAGAGAG 160

RESULT 12
US-09-242-948-3/C
? Sequence 3, Application US/99242948
? Patent No. 6252057
? GENERAL INFORMATION:
? APPLICANT: Brady, Matthew J.
```

; Printen, John A
; Saltiel, Alan R
; Warner-Lambert Company,
; (Outside USA)
; TITLE OF INVENTION: Protein Targeting to Glycogen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warner-Lambert Company
; STREET: 201 Tabor Road
; CITY: Morris Plains
; STATE: NJ
; COUNTRY: US
; ZIP: 07950
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/242,948
; FILING DATE: 25-Feb-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,107
; FILING DATE: 30-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ashbrook, Charles W
; REFERENCE/DOCKET NUMBER: 5485-01-CA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313 996-5215
; TELEFAX: 313 996-1553
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4238..5176
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-242-948-3

Query Match 0.7%; Score 54; DB 4; Length 5789;
Best Local Similarity 100.0%; Pred. No. 3.2e-12;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4389 ggaggctgaggcaggagaatcgcttgaaacccggaggcggaggttgagtgagc 4442
|||||
Db 311 GGAGGCTGAGGAGGAGGAATCGCTTGAAACCGGGGAGGAGTTGCAGTGAGC 258

RESULT 13
US-09-210-748A-3/c

; Sequence 3, Application US/09210748A
; Patent No. 6335156
; GENERAL INFORMATION:
; APPLICANT: Hermeking, Heiko
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: 14-3-3 SIGMA ARREST THE CELL CYCLE
; FILE REFERENCE: 1107.77810
; CURRENT APPLICATION NUMBER: US/09/210,748A
; CURRENT FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/069,416
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 7680
; TYPE: DNA

; ORGANISM: Homo sapiens
US-09-210-748A-3

Query Match 0.7%; Score 54; DB 4; Length 7680;
Best Local Similarity 100.0%; Pred. No. 3.1e-12;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4376 tcccagctactcgggagctgagcgagcagagaatcgcttgaaacccggagcgga 4429
|||||
Db 6086 TCCCAGCTACTCGGAGGCTGAGCGAGGAGGAATCGCTTGAAACCCGGAGCGGA 6033

RESULT 14

US-09-247-155-141
; Sequence 141, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 141
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..447
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 4..147
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.6999980926514
; OTHER INFORMATION: seq LLLFFGKLLVVG/VG
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 858..863
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 880..891
US-09-247-155-141

Query Match 0.7%; Score 52; DB 4; Length 891;
Best Local Similarity 100.0%; Pred. No. 2.2e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4375 atcccagctactcgggagctgagcgagcagagaatcgcttgaaacccggagggc 4426
|||||
Db 714 atcccagctactcgggagctgagcgagcagagaatcgcttgaaacccggagggc 765

RESULT 15

US-08-973-544-1
; Sequence 1, Application US/08973544
; Patent No. 6338950
; GENERAL INFORMATION:
; APPLICANT: WEISS, Elisabeth
; TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1

GenCore version 4.5
Copyright (c) 1995 - 2000 CompuGen Ltd.

M Nucleic - nucleic search, using sw model

Run on: July 6, 2002, 12:44:21 : Search time 16496 seconds

(without alignments)

10279,718 Million cell updates/sec

Title: US-09-761-466-4

Percent score: 78.96

Sequences: 1 cctacacccacatttcaat.....aatataaacatttattat. 7878

Search table: g1103_NUC

Gapop 60.0 , Gapext 60.0

Searched: 21979536 seqs, 1081744937 residues

Word size : 0

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 6

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cqn2_6/ptodata/2/pna/US087_COMB.seq:*

2: /cqn2_6/ptodata/2/pna/US086_COMB.seq:*

3: /cqn2_6/ptodata/2/pna/US085_COMB.seq:*

4: /cqn2_6/ptodata/2/pna/US084_COMB.seq:*

5: /cqn2_6/ptodata/2/pna/US083_COMB.seq:*

6: /cqn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cqn2_6/ptodata/2/pna/US081_COMB.seq:*

8: /cqn2_6/ptodata/2/pna/US080_COMB.seq:*

9: /cqn2_6/ptodata/2/pna/US079_COMB.seq:*

10: /cqn2_6/ptodata/2/pna/US078_COMB.seq:*

11: /cqn2_6/ptodata/2/pna/US077_COMB.seq:*

12: /cqn2_6/ptodata/2/pna/US076_COMB.seq:*

13: /cqn2_6/ptodata/2/pna/US075_COMB.seq:*

14: /cqn2_6/ptodata/2/pna/US074_COMB.seq:*

15: /cqn2_6/ptodata/2/pna/US073_COMB.seq:*

16: /cqn2_6/ptodata/2/pna/US072_COMB.seq:*

17: /cqn2_6/ptodata/2/pna/US071_COMB.seq:*

18: /cqn2_6/ptodata/2/pna/US070_COMB.seq:*

19: /cqn2_6/ptodata/2/pna/US069_COMB.seq:*

20: /cqn2_6/ptodata/2/pna/US068_COMB.seq:*

21: /cqn2_6/ptodata/2/pna/US067_COMB.seq:*

22: /cqn2_6/ptodata/2/pna/US066_COMB.seq:*

23: /cqn2_6/ptodata/2/pna/US065_COMB.seq:*

24: /cqn2_6/ptodata/2/pna/US064_COMB.seq:*

25: /cqn2_6/ptodata/2/pna/US063_COMB.seq:*

26: /cqn2_6/ptodata/2/pna/US062_COMB.seq:*

27: /cqn2_6/ptodata/2/pna/US061_COMB.seq:*

28: /cqn2_6/ptodata/2/pna/US060_COMB.seq:*

29: /cqn2_6/ptodata/2/pna/US059_COMB.seq:*

30: /cqn2_6/ptodata/2/pna/US058_COMB.seq:*

31: /cqn2_6/ptodata/2/pna/US057_COMB.seq:*

32: /cqn2_6/ptodata/2/pna/US056_COMB.seq:*

33: /cqn2_6/ptodata/2/pna/US055_COMB.seq:*

34: /cqn2_6/ptodata/2/pna/US054_COMB.seq:*

35: /cqn2_6/ptodata/2/pna/US053_COMB.seq:*

36: /cqn2_6/ptodata/2/pna/US052_COMB.seq:*

37: /cqn2_6/ptodata/2/pna/US051_COMB.seq:*

38: /cqn2_6/ptodata/2/pna/US050_COMB.seq:*

39: /cqn2_6/ptodata/2/pna/US049_COMB.seq:*

40: /cqn2_6/ptodata/2/pna/US048_COMB.seq:*

41: /cqn2_6/ptodata/2/pna/US047_COMB.seq:*

42: /cqn2_6/ptodata/2/pna/US046_COMB.seq:*

43: /cqn2_6/ptodata/2/pna/US045_COMB.seq:*

44: /cqn2_6/ptodata/2/pna/US044_COMB.seq:*

45: /cqn2_6/ptodata/2/pna/US043_COMB.seq:*

Prod. No. is the number of results produced by search. To have a score greater than or equal to the score of the result being printed, and is derived by analysis of the full score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7836	100.0	7836	1	Sequence 4, Aff 1
2	7836	100.0	7836	2	Sequence 4, Aff 1
3	1072	14.7	1072	3	Sequence 4, Aff 1
4	1072	14.7	1072	4	Sequence 4, Aff 1
5	526	6.7	1072	5	Sequence 4, Aff 1
6	489	6.2	524	6	Sequence 4, Aff 1
7	475	4.8	475	7	Sequence 4, Aff 1
8	475	4.8	475	8	Sequence 4, Aff 1
9	370	4.7	410	9	Sequence 4, Aff 1
10	329	4.2	338	10	Sequence 4, Aff 1
11	182	2.3	182	11	Sequence 4, Aff 1
12	182	2.3	182	12	Sequence 4, Aff 1
13	174	2.2	457	13	Sequence 4, Aff 1
14	98	1.3	98	14	Sequence 4, Aff 1
15	98	1.3	98	15	Sequence 4, Aff 1
16	89	1.1	89	16	Sequence 4, Aff 1
17	89	1.1	89	17	Sequence 4, Aff 1
18	89	1.1	89	18	Sequence 4, Aff 1
19	89	1.1	89	19	Sequence 4, Aff 1
20	89	1.1	89	20	Sequence 4, Aff 1
21	87	1.1	159903	21	Sequence 4, Aff 1
22	84	1.1	409	22	Sequence 4, Aff 1
23	84	1.1	1247	23	Sequence 4, Aff 1
24	84	1.1	1247	24	Sequence 4, Aff 1
25	84	1.1	1247	25	Sequence 4, Aff 1
26	84	1.1	223820	26	Sequence 4, Aff 1
27	82	1.0	898	27	Sequence 4, Aff 1
28	82	1.0	898	28	Sequence 4, Aff 1
29	81	1.0	429	29	Sequence 4, Aff 1
30	81	1.0	429	30	Sequence 4, Aff 1
31	77	1.0	1363	31	Sequence 4, Aff 1

32 77 1.0 32768 60 US-60-212-350-4 Sequence 4, Appli
33 77 1.0 32768 60 US-60-213-248-12 Sequence 12, Appl
34 77 1.0 32768 61 US-60-229-514-27 Sequence 27, Appl
c 35 77 1.0 61152 63 US-60-245-222-72 Sequence 72, Appl
c 36 77 1.0 61152 64 US-60-258-251-29 Sequence 29, Appl
c 37 1.0 124292 60 US-60-212-664-157 Sequence 157, App
c 38 77 1.0 169074 60 US-60-212-664-69 Sequence 69, Appl
c 39 77 1.0 396478 64 US-60-258-275-158 Sequence 158, App
c 40 76 1.0 8582 30 US-09-760-443-1839 Sequence 1839, Ap
c 41 76 1.0 8582 30 US-09-760-443-1990 Sequence 1990, Ap
c 42 76 1.0 8582 30 US-09-760-443-2106 Sequence 2106, Ap
c 43 76 1.0 169645 60 US-60-212-664-1 Sequence 1, Appli
c 44 76 1.0 169645 60 US-60-212-664-295 Sequence 295, App
c 45 1.0 179449 60 US-60-212-664-220 Sequence 220, App

ALIGNMENTS

RESULT 1
PCT-US01-01511-4
; Sequence 4, Application PC/TUS0101511
; GENERAL INFORMATION:
; APPLICANT: Beth Israel Deaconess Medical Center
; TITLE OF INVENTION: Cardiac-Cell Specific Enhancer Elements
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 01948/069W02
; CURRENT APPLICATION NUMBER: PCT/US01/01511
; PRIOR FILING DATE: 2000-01-16
; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(7838)
; OTHER INFORMATION: n = A,T,C or G
PCT-US01-01511-4

Query Match 100.0%; Score 7838; DB 1; Length 7838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctcgagccagaggttcaagaccagcgtggaaacatagggagagaccctctctccaca 60
Db 1 ctcgagccagaggttcaagaccagcgtggaaacatagggagagaccctctctccaca 60
QY 61 aaaaattaaaactagccaggtgtgtggcaaacacctgtagtcacagctactcagaag 120
Db 61 aaaaattaaaactagccaggtgtgtgtggcaaacacctgtagtcacagctactcagaag 120
QY 121 gctgaggtggagagatcacttgagcctggaaagttagagcgtacagtgagcgtgatcaca 180
Db 121 gctgaggtggagagatcacttgagcctggaaagttagagcgtacagtgagcgtgatcaca 180
QY 181 ccactgactccagcctgggagacagagtgagaccctgtcaataataatacaacaacaat 240
Db 181 ccactgactccagcctgggagacagagtgagaccctgtcaataataatacaacaacaat 240
QY 241 aatgattaaaataactaaaactattttatgctattttccactgtattttgtaagatt 300
Db 241 aatgattaaaataactaaaactattttatgctattttccactgtattttgtaagatt 300
QY 301 tttaaaatgaaattcccaaatgtctttccagagaggtgttccaaaattataccacatt 360
Db 301 tttaaaatgaaattcccaaatgtctttccagagaggtgttccaaaattataccacatt 360
QY 361 tcactcatgttctctctgaacagcagcaatcaggaaaaactccctggagagcgagg 420

Db 361 tcaccatgttctctctgaacagcagcaatcaggaaaaactccctggagagcgagg 420
QY 421 cttagactgagattttaaaagggttagcctcagctctctctccaggtttacactgtgc 480
Db 421 cttagactgagattttaaaagggttagcctcagctctctctccaggtttacactgtgc 480
QY 481 atgtttccaaactcaaaagattttacactctctgttgattgctctgttaagatctgc 540
Db 481 atgtttccaaactcaaaagattttacactctctgttgattgctctgttaagatctgc 540
QY 541 ccactactatgtattaaaaagggtgcatgataatgaattcagccctctctgaaaaatcc 600
Db 541 ccactactatgtattaaaaagggtgcatgataatgaattcagccctctctgaaaaatcc 600
QY 601 aaagggtcctattgcagtttcccccaatttaattgggttcattaaaaatttcttgggaaggac 660
Db 601 aaagggtcctattgcagtttcccccaatttaattgggttcattaaaaatttcttgggaaggac 660
QY 661 aaagcttttagttaactatgagaaaaaacagcagaccagccctggattctgtcttcaag 720
Db 661 aaagcttttagttaactatgagaaaaaacagcagaccagccctggattctgtcttcaag 720
QY 721 attttaccatgttgagcagcctgttagtcagagcccaagaaaaatccagccacagat 780
Db 721 attttaccatgttgagcagcctgttagtcagagcccaagaaaaatccagccacagat 780
QY 781 accctagatgtagactagcagtgctacaacctcaaggtcagagtgatgctcactagaccag 840
Db 781 accctagatgtagactagcagtgctacaacctcaaggtcagagtgatgctcactagaccag 840
QY 841 agccaaaaataggtgctatatcatttaagagagtataaaatgcaaacacagacgggtgcac 900
Db 841 agccaaaaataggtgctatatcatttaagagagtataaaatgcaaacacagacgggtgcac 900
QY 901 attattcacaataagcatataaccacaggggactcctctcgaatatgcaagaactct 960
Db 901 attattcacaataagcatataaccacaggggactcctctcgaatatgcaagaactct 960
QY 961 cactaatcaataagaaaaaggcaaaagatttaaacaggcacttcacaaaaaagatatt 1020
Db 961 cactaatcaataagaaaaaggcaaaagatttaaacaggcacttcacaaaaaagatatt 1020
QY 1021 caaaaaatcaataaacatttgaaaagatcctcaattcactagttattagggaagtgaa 1080
Db 1021 caaaaaatcaataaacatttgaaaagatcctcaattcactagttattagggaagtgaa 1080
QY 1081 ataaaccacaatgagacacccccacgccccccacgagggcgttaaaatctaaaaacatg 1140
Db 1081 ataaaccacaatgagacacccccacgccccccacgagggcgttaaaatctaaaaacatg 1140
QY 1141 taatccgaattgtttgcaaggatgcggaagaaactgccattttttgtacactgccagtatga 1200
Db 1141 taatccgaattgtttgcaaggatgcggaagaaactgccattttttgtacactgccagtatga 1200
QY 1201 gggtaaatctgtacaaaccaggttgaaaaacgctgagtagaattgactagctggtattg 1260
Db 1201 gggtaaatctgtacaaaccaggttgaaaaacgctgagtagaattgactagctggtattg 1260
QY 1261 tgaatatcatatgatccagcaatttactccttagaaatttaccacaagaaatgtgtaaa 1320
Db 1261 tgaatatcatatgatccagcaatttactccttagaaatttaccacaagaaatgtgtaaa 1320
QY 1321 catgttcacaaaagacacacgacaatctcatagagcactcactattcttaacagt 1380
Db 1321 catgttcacaaaagacacacgacaatctcatagagcactcactattcttaacagt 1380
QY 1381 caaaactggaaactacccaaatgtccatcagcagagaatggcgataaacagtagcatct 1440
Db 1381 caaaactggaaactacccaaatgtccatcagcagagaatggcgataaacagtagcatct 1440
QY 1441 tcaataatgaattgttcgacagcaatgaaagtagctacactacacaaacatgt 1500

Dd 1441 tcaataataaagtgtttccagcagcaatgaaagagtagtagtacaactacaaacaatat 1500
QY 1501 gattaaactccacaacataatactaaataatattacagacacaaaagagtgatatacta 1560
Dd 1501 gattaaactccacaacataatactaaataatattacagacacaaaagagtgatatacta 1560
QY 1561 tattatataatagtgaaactctgaaacacagcaaaaactattcttctttagaaagtcadaa 1620
Dd 1561 tattatataatagtgaaactctgaaacacagcaaaaactattcttctttagaaagtcadaa 1620
QY 1621 taattactgccttgcgggaaacagaaactcaagagagcttagtagctactagtaattgttc 1680
Dd 1621 taattactgccttgcgggaaacagaaactcaagagagcttagtagctactagtaattgttc 1680
QY 1681 tcttctcgaactgcagctagtgagcaactgttattttgtcaggtccaggtctgttcaact 1740
Dd 1681 tcttctcgaactgcagctagtgagcaactgttattttgtcaggtccaggtctgttcaact 1740
QY 1741 ggaattaaaagtcccccaaatcagaagaggttcagcaagtgaagcaagtcacactgct 1800
Dd 1741 ggaattaaaagtcccccaaatcagaagaggttcagcaagtgaagcaagtcacactgct 1800
QY 1801 ggaacttggctgggaacttagggatcccataatttgcacagccacaaagcaagcaact 1860
Dd 1801 ggaacttggctgggaacttagggatcccataatttgcacagccacaaagcaagcaact 1860
QY 1861 ttcttgcttaagtagcatctcccagatccagatcccaagatggtttgcaagcagat 1920
Dd 1861 ttcttgcttaagtagcatctcccagatccagatcccaagatggtttgcaagcagat 1920
QY 1921 gaaagcacatctcagagtagctgaaagtcttcccagttcccagttcccagttcccagttccc 1980
Dd 1921 gaaagcacatctcagagtagctgaaagtcttcccagttcccagttcccagttcccagttccc 1980
QY 1981 tctcccacttctaataaacgggcaatcacactcttcttttagaaaaaagagctgggtcccta 2040
Dd 1981 tctcccacttctaataaacgggcaatcacactcttcttttagaaaaaagagctgggtcccta 2040
QY 2041 ggcataactgtcacatctagggagagggagagagagagagagagagagagagagagag 2100
Dd 2041 ggcataactgtcacatctagggagagggagagagagagagagagagagagagagagag 2100
QY 2101 ttggtatctccaccactctgaattataaataacccagacttctctatctcccacacaca 2160
Dd 2101 ttggtatctccaccactctgaattataaataacccagacttctctatctcccacacaca 2160
QY 2161 tcttatttagaagaaagaaacgaatctcccagagctctctctaaacaaagtgttcattca 2220
Dd 2161 tcttatttagaagaaagaaacgaatctcccagagctctctctaaacaaagtgttcattca 2220
QY 2221 gaataccccctgttgagccccctggctggagagatggaaagagagagagagagagagag 2280
Dd 2221 gaataccccctgttgagccccctggctggagagatggaaagagagagagagagagagag 2280
QY 2281 aggaagatcatctccagagctaccacagactgaataactaagtcacagagtaaaccaaa 2340
Dd 2281 aggaagatcatctccagagctaccacagactgaataactaagtcacagagtaaaccaaa 2340
QY 2341 andcaaaacaaagggcatccaagcattggccacagagatggagagatgattccagttcat 2400
Dd 2341 aggcagagacaaagggcattccaagcattggccacagagatggagagatgattccagttcat 2400
QY 2401 gttctctgttccagcatcacacaggggtgcnaattgaacccatctgcaagaaacacagc 2460
Dd 2401 gttctctgttccagcatcacacaggggtgcnaattgaacccatctgcaagaaacacagc 2460
QY 2461 tagtctcccttctccacacagcaactttgttactgataataataataataactatttt 2520
Dd 2461 tagtctcccttctccacacagcaactttgttactgataataataataataactatttt 2520
QY 2521 ttttttttttaactaaaggttagataataactcaagagacacaggaataaggaagccta 2580
Dd 2521 ttttttttttaactaaaggttagataataactcaagagacacaggaataaggaagccta 2580

QY 2581 aaccacaagccttaaaataatgaaagaaagatctcatctaaagaaagctctctaaagaaag 2640
Dd 2581 aaccacaagccttaaaataatgaaagaaagatctcatctaaagaaagctctctaaagaaag 2640
QY 2641 atatttttctcccttctactttctacagtaatttttttttttttttttttttttttttttt 2700
Dd 2641 atatttttctcccttctactttctacagtaatttttttttttttttttttttttttttttt 2700
QY 2701 aaatgaaacacacagccttcccacagcacaactctatttttttttttttttttttttttt 2760
Dd 2701 aaatgaaacacacagccttcccacagcacaactctatttttttttttttttttttttttt 2760
QY 2761 cagttggaagttaaccacatctatcccacacacacacacacacacacacacacacacac 2820
Dd 2761 cagttggaagttaaccacatctatcccacacacacacacacacacacacacacacacac 2820
QY 2821 ccaagaatatttttctcctacagttctcnaaagaaagaaatcccacacacacacacacac 2880
Dd 2821 ccaagaatatttttctcctacagttctcnaaagaaagaaatcccacacacacacacacac 2880
QY 2881 actgaaacacacagaaagcctaaactccacattttatataacactctatttttttttttt 2940
Dd 2881 actgaaacacacagaaagcctaaactccacattttatataacactctatttttttttttt 2940
QY 2941 gaaagacacaaagaaatgaaatgcttttttttttttttttttttttttttttttttttttt 3000
Dd 2941 gaaagacacaaagaaatgaaatgcttttttttttttttttttttttttttttttttttttt 3000
QY 3001 aaaaactctatcagacagctttgaaatacacaaagaaagaaatgaaagaaatttttttttt 3060
Dd 3001 aaaaactctatcagacagctttgaaatacacaaagaaagaaatgaaagaaatttttttttt 3060
QY 3061 andcttgaaactctgaaagacacacagagctagagctgagtaaaagactataaacacttt 3120
Dd 3061 ggcctttgaaactctgaaagacacacagagctagagctgagtaaaagactataaacacttt 3120
QY 3121 ctgcctcgaatctcagaaagatgaaacacaaagaaagaaatcgaactaaactgaaacagac 3180
Dd 3121 ctgcctcgaatctcagaaagatgaaacacaaagaaagaaatcgaactaaactgaaacagac 3180
QY 3181 ctcttactttantaaagacacagctttgaaagagcttttttttttttttttttttttttttt 3240
Dd 3181 ctcttactttantaaagacacagctttgaaagagcttttttttttttttttttttttttttt 3240
QY 3241 cagcccccagagctcagaactgaactctgctgactttgagaaagaaagaaagaaagctagac 3300
Dd 3241 cagcccccagagctcagaactgaactctgctgactttgagaaagaaagaaagaaagctagac 3300
QY 3301 qttgcaaaactcactctctcagagtcagaaagctaaactgaaacacacacacacacacac 3360
Dd 3301 qttgcaaaactcactctctcagagtcagaaagctaaactgaaacacacacacacacacac 3360
QY 3361 ctaaaactccaaagctacactctctgaaagaaagaaagcttttttttttttttttttttttt 3420
Dd 3361 ctaaaactccaaagctacactctctgaaagaaagaaagcttttttttttttttttttttttt 3420
QY 3421 ggaagattctgctcacaactgaactctaaacaaatctgctatttttttttttttttttttttt 3480
Dd 3421 ggaagattctgctcacaactgaactctaaacaaatctgctatttttttttttttttttttttt 3480
QY 3481 aaaaaaaagaaagacactcccagaaagaaagctaaactgaacttttttttttttttttttttt 3540
Dd 3481 aaaaaaaagaaagacactcccagaaagaaagctaaactgaacttttttttttttttttttttt 3540
QY 3541 ggttcagagcagctttgaaattttgctctctcagagctgactttctaaaccccacacacac 3600
Dd 3541 ggttcagagcagctttgaaattttgctctctcagagctgactttctaaaccccacacacac 3600
QY 3601 cctctatctcagctccacacactctctctcagctccagagctcagagctcagagctcagag 3660
Dd 3601 cctctatctcagctccacacactctctctcagctccagagctcagagctcagagctcagag 3660

QY 3661 cccaaagcgggctgcctgaccccgagcgaaagggccccagtgtagtctaag 3720
|||||
Db 3661 cccaaagcgggctgcctgaccccgagcgaaagggccccagtgtagtctaag 3720
|||||
QY 3721 cgggtgcgtctctttgacagcgcggtttgggacaaacagcggggacgagagataagg 3780
|||||
Db 3721 cgggtgcgtctctttgacagcgcggtttgggacaaacagcggggacgagagataagg 3780
|||||
QY 3781 tgacataccagacagatttggtgcgcgcgtgatactcctctcccacagaaacgcgg 3840
|||||
Db 3781 tgacataccagacagatttggtgcgcgcgtgatactcctctcccacagaaacgcgg 3840
|||||
QY 3841 agctattaaaagacctatgattactttatcttcttcggaagcttcttcgagagaga 3900
|||||
Db 3841 agctattaaaagacctatgattactttatcttcttcggaagcttcttcgagagaga 3900
|||||
QY 3901 caaaagatgttccctgctaaagacacaaagccacaaacagggaggtctgcacaggcgac 3960
|||||
Db 3901 caaaagatgttccctgctaaagacacaaagccacaaacagggaggtctgcacaggcgac 3960
|||||
QY 3961 gcacaaattcggcgcggggaaagcaaaacacactgacgcttagagtgacaaaactgtgt 4020
|||||
Db 3961 gcacaaattcggcgcggggaaagcaaaacacactgacgcttagagtgacaaaactgtgt 4020
|||||
QY 4021 gtccccagacagctccagagtcggcagggacgctggggcgcgagggcgacccacag 4080
|||||
Db 4021 gtccccagacagctccagagtcggcagggacgctggggcgcgagggcgacccacag 4080
|||||
QY 4081 tatggtcttctgctccctggaaagtgtttttcacogtatgcgcgtaaaaacacgcacac 4140
|||||
Db 4081 tatggtcttctgctccctggaaagtgtttttcacogtatgcgcgtaaaaacacgcacac 4140
|||||
QY 4141 acagagaaagtgcactgagcgctgtgtgtacccgtgtgtgttttagcgaatt 4200
|||||
Db 4141 acagagaaagtgcactgagcgctgtgtgtacccgtgtgtgttttagcgaatt 4200
|||||
QY 4201 taaagcacatcaggccggggccatgctcacgcctgtaattccacagcacttttaggagcc 4260
|||||
Db 4201 taaagcacatcaggccggggccatgctcacgcctgtaattccacagcacttttaggagcc 4260
|||||
QY 4261 gaggcgccgatcacctgagtgagggttcgacacacagcctggcacaactggtgaaac 4320
|||||
Db 4261 gaggcgccgatcacctgagtgagggttcgacacacagcctggcacaactggtgaaac 4320
|||||
QY 4321 cctgtctctacaaaaatacaaaaattagccgggcatggtgatgcgtgcctgtgatccca 4380
|||||
Db 4321 cctgtctctacaaaaatacaaaaattagccgggcatggtgatgcgtgcctgtgatccca 4380
|||||
QY 4381 gctactcggaggctgaggcaggagaatcgcttgaacccggaggcgagggttcagtga 4440
|||||
Db 4381 gctactcggaggctgaggcaggagaatcgcttgaacccggaggcgagggttcagtga 4440
|||||
QY 4441 gccgagatcacaccactgcactcagcctggcgacaagacgaaattccgtctaaaaaa 4500
|||||
Db 4441 gccgagatcacaccactgcactcagcctggcgacaagacgaaattccgtctaaaaaa 4500
|||||
QY 4501 ataaaaataaaataaataaataaagccatcaactcaactcaactcaactcaactcaact 4560
|||||
Db 4501 ataaaaataaaataaataaataaagccatcaactcaactcaactcaactcaactcaact 4560
|||||
QY 4561 gttgtaagtattccatgacacaggttctaaaaatglaaacgctccattgtgctcttttaa 4620
|||||
Db 4561 gttgtaagtattccatgacacaggttctaaaaatglaaacgctccattgtgctcttttaa 4620
|||||
QY 4621 gggcttgaatgtctgcaactgtcatgtgtacacttaagtatggatgtgtcaacacgac 4680
|||||
Db 4621 gggcttgaatgtctgcaactgtcatgtgtacacttaagtatggatgtgtcaacacgac 4680
|||||
QY 4681 cctttctagcgcgtcttctgtctgtaattcccccatttccgcaatttctgtggagcg 4740
|||||
Db 4681 cctttctagcgcgtcttctgtctgtaattcccccatttccgcaatttctgtggagcg 4740
|||||
QY 4741 cagaacgcccctccggaagcgcgctgtgatcccgactttgtctccggtatcgcgcagct 4800
|||||

; PRIOR APPLICATION NUMBER: US 60/176,419
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7838
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(7838)
; OTHER INFORMATION: n = A,T,C or G
US-09-761-466-4

Query Match 100.0%; Score 7838; DB 30; Length 7838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ctcgagcccgaggttcaagaccagcctgggaacataggagagccctctctccaca 60
Db 1 ctcgagcccgaggttcaagaccagcctgggaacataggagagccctctctccaca 60
Qy 61 aaaaatttaaacactagccaggtgtggtggcaaacacactgtagtcgccgctactcagaag 120
Db 61 aaaaatttaaacactagccaggtgtggtggcaaacacactgtagtcgccgctactcagaag 120
Qy 121 gctcaggtggagatcacttgacctggagaaagtagagctacagtgagcctggtatcaca 180
Db 121 gctcaggtggagatcacttgacctggagaaagtagagctacagtgagcctggtatcaca 180
Qy 181 ccactgcactccagcctgggagacagatgagacccctgtcaaaataataacaaacaat 240
Db 181 ccactgcactccagcctgggagacagatgagacccctgtcaaaataataacaaacaat 240
Qy 241 aatgattaaataactaaactaattttatgctattttcacctgtattttgtaagatt 300
Db 241 aatgattaaataactaaactaattttatgctattttcacctgtattttgtaagatt 300
Qy 301 tttaaaatgaaaatcccaaatgtttccagaagatgttccaaatataccacatt 360
Db 301 tttaaaatgaaaatcccaaatgtttccagaagatgttccaaatataccacatt 360
Qy 361 tcaactcattctctctgaacagcagcaatcaggaaaaactccctggaagagcagg 420
Db 361 tcaactcattctctctgaacagcagcaatcaggaaaaactccctggaagagcagg 420
Qy 421 cttgagactgagattttaaaaaggggtaggctcagctcctctccaggtttacactgtgc 480
Db 421 cttgagactgagattttaaaaaggggtaggctcagctcctctccaggtttacactgtgc 480
Qy 481 atgtttccaaactcaagaattttacactctctgtgtgctattgtctgtaaaagtgcac 540
Db 481 atgtttccaaactcaagaattttacactctctgtgtgctattgtctgtaaaagtgcac 540
Qy 541 ccactactatgtattaaaaagggatgcataatgaattcagccctctctgtaaaaatcc 600
Db 541 ccactactatgtattaaaaagggatgcataatgaattcagccctctctgtaaaaatcc 600
Qy 601 aaagggctcattgagtttccccatttaattgggtcatttaaaatattcttgggaagac 660
Db 601 aaagggctcattgagtttccccatttaattgggtcatttaaaatattcttgggaagac 660
Qy 661 aaagcttttagttaactatgaaaaaacagcagaaaccagccctggattctgtcttcaag 720
Db 661 aaagcttttagttaactatgaaaaaacagcagaaaccagccctggattctgtcttcaag 720
Qy 721 attttaccatgttggcagcctgtgtccagagcccaagaataatccagccacagat 780
Db 721 attttaccatgttggcagcctgtgtccagagcccaagaataatccagccacagat 780
Qy 781 accctagatgtagactagcagtgtacaacctcaagggtcagaagtgtgtcactagaccag 840
Db 781 accctagatgtagactagcagtgtacaacctcaagggtcagaagtgtgtcactagaccag 840

Db 781 accctagatgtagactagcagtgtacaacctcaagggtcagaagtgtgtcactagaccag 840
Qy 841 agccaaaaataggtgtctatatcatattaaagagtgataaaatgcaaacacacagacaggtgac 900
Db 841 agccaaaaataggtgtctatatcatattaaagagtgataaaatgcaaacacacagacaggtgac 900
Qy 901 attattcaataaagcataataacccacacaggggactcctctatctgaatatgcataaagaactct 960
Db 901 attattcaataaagcataataacccacacaggggactcctctatctgaatatgcataaagaactct 960
Qy 961 cactaatcaataaagaaagcacaagatttaaacaggcacttccaaaaaaagtattatt 1020
Db 961 cactaatcaataaagaaagcacaagatttaaacaggcacttccaaaaaaagtattatt 1020
Qy 1021 caaaaaatcaataaacatttgaagaatcctcaattcactagttattagggaaaggtgaa 1080
Db 1021 caaaaaatcaataaacatttgaagaatcctcaattcactagttattagggaaaggtgaa 1080
Qy 1081 ataaacacacatgagacaccccccccgccccaccagacggcttaaaatcttaaacatg 1140
Db 1081 ataaacacacatgagacaccccccccgccccaccagacggcttaaaatcttaaacatg 1140
Qy 1141 taataccgaattttgcaaggatgagaaactgccattttttgtacactgccagtatga 1200
Db 1141 taataccgaattttgcaaggatgagaaactgccattttttgtacactgccagtatga 1200
Qy 1201 gggtaaatctgtacaaccaggttggaaaacgctgagtagaatttactctagctggatttg 1260
Db 1201 gggtaaatctgtacaaccaggttggaaaacgctgagtagaatttactctagctggatttg 1260
Qy 1261 tgaatatcatatgatccagcaatttctactcctagaaatttaccocacagaaaaatgtgtaaa 1320
Db 1261 tgaatatcatatgatccagcaatttctactcctagaaatttaccocacagaaaaatgtgtaaa 1320
Qy 1321 catgttccacaaaagacacacgaacaaattcatagaggcactcactattcttaacagt 1380
Db 1321 catgttccacaaaagacacacgaacaaattcatagaggcactcactattcttaacagt 1380
Qy 1381 caaaaactggaaactaccacaaatgtccatcagcagagaaatggcgataaacagtagcatct 1440
Db 1381 caaaaactggaaactaccacaaatgtccatcagcagagaaatggcgataaacagtagcatct 1440
Qy 1441 tcacataatgaatgtttcgacagcaatgaaaagttagctagctacaaactacaacaatgt 1500
Db 1441 tcacataatgaatgtttcgacagcaatgaaaagttagctagctacaaactacaacaatgt 1500
Qy 1501 gattgaacctcacaaacataatactaaagtaaaattatcagacacaaagagtgtatatactg 1560
Db 1501 gattgaacctcacaaacataatactaaagtaaaattatcagacacaaagagtgtatatactg 1560
Qy 1561 tatttagatcatgtgaagtctgaaacacagggcaaaaactattctgtttagaagtcagaa 1620
Db 1561 tatttagatcatgtgaagtctgaaacacagggcaaaaactattctgtttagaagtcagaa 1620
Qy 1621 tagttactccctgcgggaaacacagaaactcaagagggttagtagctactcgttaagtcttc 1680
Db 1621 tagttactccctgcgggaaacacagaaactcaagagggttagtagctactcgttaagtcttc 1680
Qy 1681 tgcctcctgaactgcattgcttagtgaggcagctgtattttgtgcagtcctgtgttacact 1740
Db 1681 tgcctcctgaactgcattgcttagtgaggcagctgtattttgtgcagtcctgtgttacact 1740
Qy 1741 ggagttaaaagtccccccaaaaatcagaaggtgttcagcaagtggaagcaagtagcacctgct 1800
Db 1741 ggagttaaaagtccccccaaaaatcagaaggtgttcagcaagtggaagcaagtagcacctgct 1800
Qy 1801 ggaacttggctgggaacttaggggatcccaataatttgcacaggcacaagcaagccaagct 1860
Db 1801 ggaacttggctgggaacttaggggatcccaataatttgcacaggcacaagcaagccaagct 1860
Qy 1861 ttcctgccttaagtagcatctccagagtcaggatccaggaaatggtttggcaggcaggaat 1920
Db 1861 ttcctgccttaagtagcatctccagagtcaggatccaggaaatggtttggcaggcaggaat 1920

4081	Db	 tatggtctctgtgcccccttggaagatttttttcccggtatgcggtataaacacgcacac	4140
4141	QY	acagagaagtgactgtcacttaaggcgccgtgtgtgtaccogtgcgttttagcgaatt	4200
4141	Db	acagagaagtgactgtgtcacttaaggcgccgtgtgtgtaccogtgcgttttagcgaatt	4200
4201	QY	taagcacactcagcgccggcgccatgggtcaccgctgtaatcccagcacttttagaggcc	4260
4201	Db	taagcacactcagcgccggcgccatgggtcacgctgtaatcccagcacttttagaggcc	4260
4261	QY	gaggcgggccgatcaacctgaggtcgggagttcgacacacagcctggccaaacatggtgaaac	4320
4261	Db	gaggcgggccgatcaacctgaggtcgggagttcgacacacagcctggccaaacatggtgaaac	4320
4321	QY	cctgtctctacaaaaatacaaaaatagccggcgcatgggtgatgcgtgcctgtgatccca	4380
4321	Db	cctgtctctacaaaaatacaaaaatagccggcgcatgggtgatgcgtgcctgtgatccca	4380
4381	QY	gtactcgggagctgagcgaggagaaatcgcttgaacccgggagcgaggttcagtgga	4440
4381	Db	gtactcgggagctgagcgaggagaaatcgcttgaacccgggagcgaggttcagtgga	4440
4441	QY	gccgaagtacacacactcactccagcctggcgacagagcgaataccgctctaaaaaa	4500
4441	Db	gccgaagtacacacactcactcccgccctggcgacagagcgaataccgctctaaaaaa	4500
4501	QY	ataaaataaaataaaatgataatgaagccatcaactcaattcaaaagcgttactggtg	4560
4501	Db	ataaaataaaataaaatgataatgaagccatcaactcaattcaaaagcgttactggtg	4560
4561	QY	gttgaatgtatccatagacacagctctaaaaatgtataacgctccattgtgctccttttaa	4620
4561	Db	gttgaatgtatccatagacacagctctaaaaatgtataacgctccattgtgctccttttaa	4620
4621	QY	gggcttgaatgtctgaactgtcatgtgtacacttaaaatgtggatgtgtcaaacacgac	4680
4621	Db	gggcttgaatgtctgaactgtcatgtgtacacttaaaatgtggatgtgtcaaacacgac	4680
4681	QY	cccttctagcgcgctggtttcggtctgaatccccgcgatttcgcgaattgtgtggagcg	4740
4681	Db	cccttctagcgcgctggtttcggtctgaatccccgcgatttcgcgaattgtgtggagcg	4740
4741	QY	cagaaagccctccgcgaagcgcgctgtgtatcccgcactttgctcgggtatcgcgagct	4800
4741	Db	cagaaagccctccgcgaagcgcgctgtgtatcccgcactttgctcgggtatcgcgagct	4800
4801	QY	tgttggtcctccgggtcccccggtgccatgccccgggaggtctccacagacacgcgttcg	4860
4801	Db	tgttggtcctccgggtcccccggtgccatgccccgggaggtctccacagacacgcgttcg	4860
4861	QY	gccgaattatacagacactgaattgggttttttttggtgtgtgtgtgcaacaacaattgt	4920
4861	Db	gccgaattatacagacactgaattgggttttttttggtgtgtgtgtgcaacaacaattgt	4920
4921	QY	cagctgtgtttacaatgcgtcccgcgcggtggtaaaccttggctgcggttaacgcacag	4980
4921	Db	cagctgtgtttacaatgcgtcccgcgcggtggtaaaccttggctgcggttaacgcacag	4980
4981	QY	caggttggagggcacgacccggaagaaagagcgagggagggaaagcgcgacacct	5040
4981	Db	caggttggagggcacgacccggaagaaagagcgagggagggaaagcgcgacacct	5040
5041	QY	agggcccgctggcgacgcgtttccagcatcaattcagcactgagccggcgacgacgaca	5100
5041	Db	agggcccgctggcgacgcgtttccagcatcaattcagcactgagccggcgacgacgaca	5100
5101	QY	gggctggggggtccccggaagttcggccagccggggttttgggccagagcgcgaggtctgc	5160
5101	Db	gggctggggggtccccggaagttcggccagccggggttttgggccagagcgcgaggtctgc	5160
5161	QY	ccggtgggttaggtgcgactcttcacctctccggggagcgcgccgcgacgaccccaaccac	5220

Db	5161	cgctgtgttagtgcgactctctcaactctccggggcgccgacgaccccaaccaccc	52220
Qy	5221	cgcaagcgtcgctcgcccgctgtgtcccccgcgcgggcacaaaaacagcgcgagtt	52800
Db	5221	cgcaagcgtcgctcgcccgctgtgtcccccgcgcgggcacaaaaacagcgcgagtt	52800
Qy	5281	cgccagctctctttcccaaaacctgaacgccaaagcgcgaaggttttccaaagtcgggt	53400
Db	5281	cgccagctctctttcccaaaacctgaacgccaaagcgcgaaggttttccaaagtcgggt	53400
Qy	5341	tcccgcggttccacaccgcgcggcagcgcgaaacagcccaggagcaaccatttcttc	54000
Db	5341	tcccgcggttccacaccgcgcggcagcgcgaaacagcccaggagcaaccatttcttc	54000
Qy	5401	ttcactgtatctgagtcgtgttcctatctgactcgaatgtcaactgatattccagctgtg	54600
Db	5401	ttcactgtatctgagtcgtgttcctatctgactcgaatgtcaactgatattccagctgtg	54600
Qy	5461	acctccagcaggaactccgaggaactgaattccagctctcgattctctccgcctctcc	55200
Db	5461	acctccagcaggaactccgaggaactgaattccagctctcgattctctccgcctctcc	55200
Qy	5521	gcccgttttgctgaagcgttttcagccgttcgggcagaaaggggtgggatgtggcagcc	55800
Db	5521	gcccgttttgctgaagcgttttcagccgttcgggcagaaaggggtgggatgtggcagcc	55800
Qy	5581	accagcccccagcagagaagaaaagaggagcaaaattaacgcgaaaggacacccggaagtc	56400
Db	5581	accagcccccagcagagaagaaaagaggagcaaaattaacgcgaaaggacacccggaagtc	56400
Qy	5641	tgaagcgaactccctcgatcctcgaaatccgagcgcaaaccttaacactagtttgaagc	57000
Db	5641	tgaagcgaactccctcgatcctcgaaatccgagcgcaaaccttaacactagtttgaagc	57000
Qy	5701	ggatcatatccactaatccaggacaaaattcgggttgggaaaaatactccccagagcctaa	57600
Db	5701	ggatcatatccactaatccaggacaaaattcgggttgggaaaaatactccccagagcctaa	57600
Qy	5761	gaaactgacttacaacaaacaaactgacaagacaaaaatgcgaagagtttgtgaaa	58200
Db	5761	gaaactgacttacaacaaacaaactgacaagacaaaaatgcgaagagtttgtgaaa	58200
Qy	5821	cgtaatgtctcagaanaatgtgtatataatacatcctataataatgttttaaatttg	58800
Db	5821	cgtaatgtctcagaanaatgtgtatataatacatcctataataatgttttaaatttg	58800
Qy	5881	caaaaaaaagtcctaaagagatataattttaaaaaccagttggcacttgggaggaagt	59400
Db	5881	caaaaaaaagtcctaaagagatataattttaaaaaccagttggcacttgggaggaagt	59400
Qy	5941	gggattagctgagaaggggagaaggaagcatttttgagtgagcgtaaatgttttgtatc	60000
Db	5941	gggattagctgagaaggggagaaggaagcatttttgagtgagcgtaaatgttttgtatc	60000
Qy	6001	ttgatatgttggtgtgttatggggtgcacatcccaagttcgaagactatcgaactgtac	60600
Db	6001	ttgatatgttggtgtgttatggggtgcacatcccaagttcgaagactatcgaactgtac	60600
Qy	6061	actttgtctaggctacattagacctcaataaagtggattttaaacctaaataagccagg	61200
Db	6061	actttgttctaggctacattagacctcaataaagtggattttaaacctaaataagccagg	61200
Qy	6121	taacagctttgcctgggtggctgggggagagcgttgggacactttacattgatctccctc	61800
Db	6121	taacagctttgcctgggtggctgggggagagcgttgggacactttacattgatctccctc	61800
Qy	6181	ttagcatgttcggtttgtgtgtgtgtgtttcttatgatgtattattatccaaaaatat	62400
Db	6181	ttagcatgttcggtttgtgtgtgtgtgtttcttatgatgtattattatccaaaaatat	62400
Qy	6241	atcattagcagagtactgatgtcaaatgtaaaaaccattgttaaggaaacacaaagcgc	63000
Db	6241	atcattagcagagtactgatgtcaaatgtaaaaaccattgttaaggaaacacaaagcgc	63000


```
Query Match 4.7% Score 370; DB 25; Length 410;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4749 cttccgcaaaagcgctgtatcccaacttttctccatatacgcacacttggttgcac 4808
Db 5 cttccgcaaaagcgctgtatcccaacttttctccatatacgcacacttggttgcac 68
QY 4809 tccggttccccatccatgcccccggagagctctccaaacacacgttcgcccaatt 4868
Db 6 tccggttccccatccatgcccccggagagctctccaaacacacgttcgcccaatt 128
QY 4869 atacaaactaaatgggtttttttatgtatgtatgtatgtatgtatgtatgtatgt 4928
Db 129 atacaaactaaatgggtttttttatgtatgtatgtatgtatgtatgtatgtatgt 188
QY 4929 gtccaaatgcctcccgccggcggtgaaacttgcctgcatacgcacagcagttgc 4988
Db 189 gtccaaatgcctcccgccggcggtgaaacttgcctgcatacgcacagcagttgc 248
QY 4989 aggcacacacccggaaggaaagaaacacacacacacacacacacacacacacacac 5048
Db 249 aggcacacacccggaaggaaagaaacacacacacacacacacacacacacacacac 308
QY 5049 tgcacacacgtttccacacatcaattcaactgaacccgcgcgcgcgcgcgcgcgcgc 5108
Db 309 tgcacacacgtttccacacatcaattcaactgaacccgcgcgcgcgcgcgcgcgcgc 368
QY 5109 gactcccccga 5118
Db 369 gactcccccga 378

RESULT 10
US-60-182-316-1254
: Sequence 1254, Application US/60182316
: GENERAL INFORMATION:
: APPLICANT: Curtis, Anne L.
: APPLICANT: Laquae, Robert E.
: APPLICANT: Klingler, Tod M.
: APPLICANT: Stuve, Laura L.
: TITLE OF INVENTION: CpG Island Polynucleotides
: FILE REFERENCE: PX-0003 P
: CURRENT APPLICATION NUMBER: US/60/182,316
: CURRENT FILING DATE: 2000-02-10
: NUMBER OF SEQ. ID NOS: 14,630
: SOFTWARE: PERL Program
: SEQ. ID NO 1254
: LENGTH: 338
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: OTHER INFORMATION: Incyte ID No: CpG_991027_R15_masked_fa.Contin29558
US-60-182-316-1254

Query Match 4.2% Score 329; DB 57; Length 338;
Best Local Similarity 100.0%; Pred. No. 4.1e-90;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3883 aactctcttcggaagacacaaataatctccatccctaaacacacacacacacacacac 3942
Db 10 aactctcttcggaagacacaaataatctccatccctaaacacacacacacacacacac 69
QY 3943 aactctcttcggaagacacacaaatctccgctggggaagaaacacacacacacacacac 4002
Db 70 aactctcttcggaagacacacaaatctccgctggggaagaaacacacacacacacacac 129
QY 4003 gaatcacacacacacacacacacacacacacacacacacacacacacacacacacac 4062
Db 330 gaatcacacacacacacacacacacacacacacacacacacacacacacacacacac 189
```

```
QY 4063 gacaaagcgacacacacacacacacacacacacacacacacacacacacacacac 4122
Db 190 gacaaagcgacacacacacacacacacacacacacacacacacacacacacacac 249
QY 4123 gacaaagcgacacacacacacacacacacacacacacacacacacacacacacac 4182
Db 250 gacaaagcgacacacacacacacacacacacacacacacacacacacacacacac 466
QY 4183 gacaaagcgacacacacacacacacacacacacacacacacacacacacacacac 4241
Db 310 gacaaagcgacacacacacacacacacacacacacacacacacacacacacacac 438

RESULT 11
US-08-369-881-1374/c
: Sequence 1374, Application US/08369881
: GENERAL INFORMATION:
: APPLICANT: Wilde, Craig S.
: APPLICANT: Deleqane, Angelo M.
: APPLICANT: Mills, Pamela Kay
: APPLICANT: Pham, Miro Thoi
: TITLE OF INVENTION: HUMAN CARDIAC CELL-DEVELOPMENTAL PROTEIN ENZYMES
: TITLE OF INVENTION: AND POLYPEPTIDES
: NUMBER OF SEQUENCES: 3690
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3330 Hillview Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Word Perfect 6.0/6.1/MS-WS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/369,881
: FILING DATE: Herewith
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Barbara J. Luther
: REGISTRATION NUMBER: 43,954
: REFERENCE/DOCKET NUMBER: P00019 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-855-0572
: INFORMATION FOR SEQ ID NO: 1374:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 182 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: IMMEDIATE SOURCE:
: CLONE: 185540
US-08-369-881-1374

Query Match 2.3% Score 182; DB 7; Length 182;
Best Local Similarity 100.0%; Pred. No. 4.1e-47;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2781 taccctccacacacacacacacacacacacacacacacacacacacacacacac 2840
Db 182 taccctccacacacacacacacacacacacacacacacacacacacacacacac 123
QY 2841 taccctccacacacacacacacacacacacacacacacacacacacacacacac 2900
Db 122 taccctccacacacacacacacacacacacacacacacacacacacacacacac 43
QY 2901 taccctccacacacacacacacacacacacacacacacacacacacacacacac 2960
```

```
|||||
Db 62 GACTCCCATGTTTAGAGGTCATTGATGGGTTTGTGCATGAAGGCAGGAGGACTGA 3
QY 2961 ga 2962
|||
Db 2 GA 1

RESULT 12
US-08-408-872-494/c
; Sequence 494, Application US/08408872
; GENERAL INFORMATION:
; APPLICANT: Wilde, Craig G.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Bills, Pamela K.
; APPLICANT: Pham, Mino T.
; APPLICANT: Altus, Christina M.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: CARDIAC CELL-DERIVED POLYNUCLEOTIDES AND
; NUMBER OF SEQUENCES: 1954
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3330 HILLVIEW AVENUE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,872
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LUTHER, BARBARA J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PD-0032 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555
; TELEFAX: (415) 852-0195
; INFORMATION FOR SEQ ID NO: 494:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 182 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; IMMEDIATE SOURCE:
; CLONE: 185540
US-08-408-872-494

Query Match 2.3%; Score 182; DB 8; Length 182;
Best Local Similarity 100.0%; Pred. No. 3.1e-45;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2781 gatcccggaagtcggaaggagcaagtcgtgtgaagccaagaggtatctttcccta 2840
Db 182 GATCCCGGGAAGTCGGAAGGAGGACAGTCGTGTGAAGCCAAGAGGTATCTTTCCCTA 123
QY 2841 cagcttctcaagagagggagcccggtgggtaattglagcgtggaacacccagagcgt 2900
Db 122 CAGCTTCTCAAGAGAGGGGATCCCGGTGGTAAATGTGAGGCTGGAACACCCGAGAGCT 63
QY 2901 gactcccatgtttatagaggtcattgatgggtttgtcatggaaggcaggaggagactga 2960
Db 62 GACTCCCATGTTTAGAGGTCATTGATGGGTTTGTGCATGAAGGCAGGAGGACTGA 3
QY 2961 ga 2962
```

```
|||
Db 2 GA 1

RESULT 13
US-09-866-555-406/c
; Sequence 406, Application US/09866555
; GENERAL INFORMATION:
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Algarte, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Klee, Jennifer
; TITLE OF INVENTION: LUNG TUMOR-ASSOCIATED SEQUENCES AND METHODS OF USE
; FILE REFERENCE: 210121.580
; CURRENT APPLICATION NUMBER: US/09/866,555
; CURRENT FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 20487
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 406
; LENGTH: 457
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(457)
; OTHER INFORMATION: n = A,T,C or G
US-09-866-555-406

Query Match 2.2%; Score 174; DB 33; Length 457;
Best Local Similarity 99.4%; Pred. No. 7.9e-43;
Matches 344; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 6033 coaagtgcaagcaactcagaactatatacactttgttctaggtacattagacctcaataa 6092
Db 351 CCAAGTGCAAGACTCATCGAACTGTACACTTTTGTCTAGGTACATTAGACCTCAATAA 292
QY 6093 agtggattttaacctaaataagccagggtaacagacttgcctgggtggctgggggagagg 6152
Db 291 AGTGGATTTTAAACCTGAATAAGCCAGGTAACAGACTTTGCCCTGGGTGGCGGAGAGG 232
QY 6153 ctggggacaacttaccattgatccctctctctctctctctctctctctctctctctctc 6212
Db 231 CTTGGGACACTTTACATTGATCTCCCTCTTAGGCATGTTCTGTTTGGTTTGGTTTCTTC 172
QY 6213 ttatgatattattatttttttttttttttttttttttttttttttttttttttttttt 6271
Db 171 TTATGATGTTATTATTATTTCGAAATAATATCATTAGCAGAGTGACTGATGTAATGTAA 112
QY 6272 aaccattgttaaggaaacccaacaaagcgggaacagagacacactggtgcacacctgttaga 6331
Db 111 AACCATTTGTAAGGAACCAACAAAGCGGGAACAGACACACTGCTGTCATCTCTGTAGA 52
QY 6332 gggataagaataagcactcgtgtccaaagctcataaaattttgg 6377
Db 51 GGGATAAGAATAAGCACTCGCTGTCCAGCTCATATAAATAATTTTGG 6

RESULT 14
PCT-US01-01354-41534
; Sequence 41534, Application PC/TUS0101354
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC004PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01354
; CURRENT FILING DATE: 2001-03-17
; NUMBER OF SEQ ID NOS: 42506
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41534
; LENGTH: 8965
```


; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,065
; PRIOR FILING DATE: 2000-09-14

; PRIOR APPLICATION NUMBER: 60/232,398
Query Match 1.3%; Score 98; DB 30; Length 8965;
Best Local Similarity 100.0%; Pred. No. 1e-19;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4375 atcccagctactcgggaggtgagcgagaggaatcgctgaacccgggagcgaggttg 4434
||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1283 atcccagctactcgggaggtgagcgagaggaatcgctgaacccgggagcgaggttg 1342
||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 4435 cagtgagcgagatcacaccactgcactcagcctggg 4472
||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1343 cagtgagcgagatcacaccactgcactcagcctggg 1380

Search completed: July 7, 2002, 01:14:25
Job time: 45064 sec

Sequence version 4.5
Copyright (c) 1993 - 2000 Compuen Ltd.

CM nucleic - multiple search, using sw model

Run on: July 6, 2002, 02:15:11 : Search time 10026.1 seconds
(without alignments)
10551.496 Million cell updates/sec

Title: US-09-761-466-4

Partner score: 748

Sequence: 1 ctcacccacccagattctatc.....aataataaacgataaanaa 7848

Scoring table: 4 LIGI_NUC

Gapop 60.0 , Gapext 60.0

Searched: 14736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2090000060

Post processing: Listing first 45 summaries

Database :

EST: *

- 1: em_estba: *
- 2: em_esthum: *
- 3: em_estlin: *
- 4: em_estmu: *
- 5: em_estov: *
- 6: em_estpl: *
- 7: em_estro: *
- 8: em_hcci: *
- 9: qb_estli: *
- 10: qb_est2: *
- 11: qb_hic: *
- 12: qb_oss: *
- 13: em_oss_hum: *
- 14: em_oss_inv: *
- 15: em_oss_fla: *
- 16: em_oss_vrt: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	415	5.3	641	19	BIR35729
2	401	5.1	451	9	A194753
3	252	3.2	308	9	AA994279
4	196	2.5	493	10	BF108955
5	174	2.2	457	10	N69866
6	95	1.1	536	12	AQ262855
7	82	1.0	494	10	N76766
8	81	1.0	423	9	AA224595
9	81	1.0	482	9	AA225044
10	77	1.0	376	9	AV712092
11	75	1.0	453	10	BE100645
12	75	1.0	526	9	AW957414
13	75	1.0	569	9	AA527961
14	75	1.0	2820	11	AF118360
15	74	0.9	242	9	BE051830
16	74	0.9	358	9	A1054827
17	74	0.9	441	12	A0818830

C 18	74	0.9	489	9	AK072104
C 19	74	0.9	670	12	A0145449
C 20	73	0.9	544	10	B041474
C 21	73	0.9	683	12	A0145449
C 22	72	0.9	643	9	A1241444
C 23	72	0.9	643	12	A1241444
C 24	71	0.9	454	10	A1241444
C 25	71	0.9	206	10	B051471
C 26	70	0.9	341	10	B081411
C 27	69	0.9	145	16	B056844
C 28	68	0.9	206	10	B051471
C 29	68	0.9	341	10	B081411
C 30	68	0.9	412	9	A1607412
C 31	68	0.9	443	10	N241444
C 32	68	0.9	443	9	AW80764
C 33	68	0.9	443	9	AW80764
C 34	68	0.9	444	12	AG47747
C 35	68	0.9	517	10	B549477
C 36	68	0.9	518	12	AG17847
C 37	68	0.9	518	12	B60004
C 38	67	0.9	442	9	AB27244
C 39	67	0.9	554	9	AF57644
C 40	66	0.8	407	10	AB57644
C 41	66	0.8	448	10	N68478
C 42	66	0.8	547	12	AG17111
C 43	66	0.8	571	9	AA57244
C 44	66	0.8	582	10	BF22444
C 45	65	0.8	592	9	A343544

BI035729	641 bp	cDNA	linear	EST	4-10-12
6635653761 NIH_MGC_120 Homo sapiens cDNA clone IMAGE522445					
mRNA sequence:					
BI035729	6415729.1	61:15947279			
EST.					
human.					
Homo sapiens					
Fukuyama, Motomizu, Chikudate, Graciatia, Vortubata; Euteleostomus					
Mammalia; Metazoa; Primates; Carcharia; Mammalia; Gene Collection (MGC)					
1 (bases 1 to 641)					
NIH-MGC help://mgc.col.nih.gov/					
National Institutes of Health, Mammalian Gene Collection (MGC)					
Unpublished (1999)					
Contact: Robert Strausfeld, Ph.D.					
Email: rstrausf@mail.nih.gov					
Tissue Procurement: Life Technologies, Inc.					
cDNA Library Preparation: Life Technologies, Inc.					
cDNA Sequencing: NIH-MGC, M.A.B. Laboratory (LNL)					
Clone distribution: Mammalian Gene Collection					
Found through the L.M.A.B. Laboratory information on the					
http://lmlab.col.nih.gov					
Plate: L.M.A.B. Laboratory					
High quality sequence stop: 639.					
Location/Qualifiers					
1. 641					
/Database "BIR35729"					
/Accession "BIR35729.1"					
/Name "IMAGE522445"					
/Accession "IMAGE522445"					
/Date "1999-07-12"					
/Lab "BIR35729"					
/Note "cDNA clone isolated from NIH-MGC library"					
/Source "NIH-MGC, M.A.B. Laboratory (LNL)"					
/Source anonymous pool of spleen and pancreas from 26 yr male. Library is oriented primed and directionally cloned (BIR35729 site is 3'-terminal open reading frame). Average insert size 1.5 kb. Insert size range 1.2-2.5 kb. Library is					

normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH_MGC Library."

BASE COUNT 114 a 210 c 204 g 113 t
ORIGIN

Query Match 5.3%; Score 415; DB 10; Length 641;
Best Local Similarity 100.0%; Pred. No. 3.8e-109;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4704 gctgaatccccgcatttcgccaatttgccttgagcgcagaaacccctccgcgaagcgc 4763
Db 1 GTCTGAATCCCGCATTTGCGCAATTTGCTTGGAGCGCAGAACGCCCTCCGCAAGCGC 60

QY 4764 qctgctatccccgacttgcctcggtatccgcagcttgcctgcctcggtcccccgtg 4823

Db 61 CTGCTGATCCCGACTTGTCTCCGGTATCGGCAGCTGTGTGGCTCGGGTCCCGCGTG 120

QY 4824 ccatgccccggaggctctccacagacacgccttgcgcgaattatcacgagactgaatg 4883

Db 121 CCATGCCCGGGAGGCTCTCCACAGACACCGCTTGGCGCGAATTATACGAGACTGAATG 180

QY 4884 ggtttttttggtgtgtgtgcaacacacaaatttgcagctgtgttcaaatgcgtc 4943

Db 181 GGTTTTTTGGTGTGTGTGCAACACAAATTTGTGAGTGTCTGTCAAAATCGCGTC 240

QY 4944 cgcgcgcgtggaacttgctgcgtacgcacagcagcttgagggcacgaccgcga 5003

Db 241 CGCCGGCGGTGGAAACTTGGCTCGGTACGCACAGCAGGTTGGAGGGCACGACCCGGA 300

QY 5004 aggaagaagcgcagagaggaagcgcgcacctaggccgcctggccgcgctttcc 5063

Db 301 AGGAAGAGAGGCGAGGAGGAAAGCGGCGACCTTAGGCCCTGGCAGCGCTTCC 360

QY 5064 agcatatcagcactgagccgcgcagcagcagcagcagcagcagcagcagcagcagc 5118

Db 361 AGCATCAATTCAGCACTGAGCGCGCGCAGCAGCAGCAGCAGCGGCTGGGGGCTCCCGGA 415

RESULT 2
AI393753/c
LOCUS AI393753 461 bp mRNA linear EST 04-FEB-1999
DEFINITION tg60d12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2113175 3', mRNA sequence.

ACCESSION AI393753
VERSION AI393753.1 GI:4223300
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 461)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-rc@mail.nih.gov

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 444.

FEATURES
source
1..461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2113175"
/lab_host="Soares_NSF_F8_9W_OT_PA_P_S1"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized

libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and cloneIDs: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HP8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 139 a 118 c 67 g 137 t
ORIGIN

Query Match 5.1%; Score 401; DB 9; Length 461;
Best Local Similarity 99.8%; Pred. No. 5.4e-105;
Matches 451; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5926 gcttggggaggagtgaggattagctgagaaaggggagaaagcatttttgagggtgacgt 5985

Db 461 GCTTGGGAGGAGTGGGATTAGCTGAGAGGGGAGAGGAGCATTTTGTAGGTGACGT 402

QY 5986 aaatgtttttgattctgattatggtgctgttatgggggtgcacatccaaagtgtcaaga 6045

Db 401 AAATGTTTTTGTATCTTGATTATGCTGGCTGTATGGGGGTGCACATCCAAGTGTCAAGA 342

QY 6046 ctcatcgaaactgtaacatttcttaggtacatagacctcaataaaatggatttttaa 6105

Db 341 CTCATCGAACTGTACACTTTTGTCTAGGTACATTAGACCTCAATAAAGTAGATTTTAA 282

QY 6106 cctaaataagccagtaaacagctttgcctggctggcgggagagcgttgggacacttt 6165

Db 281 CCTAATAAGCCAGGTAAACAGCTTTGCCCTGGGTGGCTGGGGGAGAGCGCTTGGGACACTTT 222

QY 6166 acattgatctccctcttaggcattgttcgttttggtttggtttcttcttatgatgtatta 6225

Db 221 ACATTGATCTCCCTCTTAGGCATGTTGCTTTTGGTTTGGTTTCTTCTATGATGTATTA 162

QY 6226 ttattcctaaataatcatatagcagagtgactgatgataaaatgtaaaccaattgttaag 6285

Db 161 TTTATTTCAAAAATATATCATTAGCAGAGTACTGATGTAATGTAATAACCATTTGTTAAG 102

QY 6286 aaaccaacaagaacggggaacagacactggtgcctctgttagagggtataagaataag 6345

Db 101 AAACCAACAAGCGGGGAACAGACACTGCTGTCCTCTGTAGAGGGGATAGGATAG 42

QY 6346 cactcgtgtccagctcataaaatatttgg 6377

Db 41 CACTCGCTGTCCAAGCTCATAAATATTTTGG 10

RESULT 3
AA994279/c
LOCUS AA994279 308 bp mRNA linear EST 27-AUG-1998
DEFINITION ou05f09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1625417 3', mRNA sequence.

ACCESSION AA994279
VERSION AA994279.1 GI:3180824
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 308)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-rc@mail.nih.gov

M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

TITLE
JOURNAL
COMMENT

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ml3 -40 forward
High quality sequence stop: 298.

FEATURES

source

1. .457
/organism="Homo sapiens"
/db_xref="GDB:1242571"
/db_xref="taxon:9606"
/clone="IMAGE:297649"
/clone_lib="Soares_fetal_lung_NbHL19w"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: p7T13D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCATCTGAAGTGGGAGCGCCGAATTTTCTTTT-3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T13 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19w."

BASE COUNT 133 a 119 c 69 g 132 t 4 others
ORIGIN

Query Match 2.2%; Score 174; DB 10; Length 457;

Best Local Similarity 99.4%; Pred. No. 4.6e-40;

Matches 344; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 6033 ccaagtgtcaagactcatcgaaactgtacacatttggcttaggtacattagaccctcaataa 6092

Db 351 CCAAGTGTCAAGACTCATCGAACTGTACACTTTGTTCTAGGTACATTAGACCTCAATAA 292

Qy 6093 agtggattttaaactaaataagccaggtaaacagcttgcctgggtgctggggagagg 6152

Db 291 AGTGGATTTTAAAGCTGAATAAGCCAGGTAACAGCTTTGCCGTGGGTGGCTGGGGGAGAGG 232

Qy 6153 cttgggacactttacattgatctccctcttagcatgtctcttggttggttggttgttc 6212

Db 231 CTTGGGACACTTTACATTGATCTCCCTCTTAGGCATGTCTGTTGGTTGGTTTGTTC 172

Qy 6213 ttatgatgtattatttttc-aaaaatatatcattagcagagtgtactgatgtataatgtaa 6271

Db 171 TTATGATGTATTATTATTCGAAAAATATATCATTTAGCAGAGTGACTGATGTAATGTAA 112

Qy 6272 aaccattgttaagaaacccaacaaagcgggaacaaagagacactggtgcactcctgttaga 6331

Db 111 AACCATTTTAAAGAAACCAACAAAGCGGGAACAAAGAGACACTGGTGCATCTCTTTAGA 52

Qy 6332 gggataagaataagcactcgcgtccaaagctcataaaatttgg 6377

Db 51 GGGATAAGAATAAGCACTCGCTGTGTCCAAGCTCATATAATATTTTGG 6

RESULT 6

AQ262855

LOCUS

DEFINITION CITBI-El-2511A10.TR CITBI-El Homo sapiens genomic clone 2511A10,

536 bp DNA linear GSS 24-OCT-1998

ACCSSION

AQ262855

VERSION AQ262855.1 GI:3789335

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 536)

AUTHORS

Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venter, J.C.

TITLE

Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building

JOURNAL

Unpublished (1998)

COMMENT

Other_GSSs: CITBI-El-2511A10.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse

Class: BAC ends.

FEATURES

source

1. .536

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="2511A10"

/clone_lib="CITBI-El"

/sex="male"

/cell_type="sperm"

/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;

CalTech Human BAC Library D'

BASE COUNT 166 a 121 c 131 g 118 t

ORIGIN

Query Match 1.1%; Score 85; DB 12; Length 536;

Best Local Similarity 100.0%; Pred. No. 1.1e-14;

Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4395 tgaggcaggagaatcgcttgaaaccggagcgagggttgagtgagcgagatcacacac 4454

Db 108 TGAGCCAGAGAAATCGCTTGAACCCGGGAGCGGAGTTGCAGTGAGCCGAGATCACACC 167

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Qy 4455 actgcactccagcctggcgacaag 4479

Db 168 ACTGCACCTCCAGCCTGGGCGACAAG 192

Stefanienberg, K., Soares M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, K., Weidmann, P., Watanabe, R., Wilson, K. and Warrick, M. Generation and analysis of 260,000 human expressed sequence tags

JOURNAL
MEDLINE
COMMENT

Genome Res. 6 (4): 807-828 (1996)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8601, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est.watson.wustl.edu

This clone is available royalty-free through INML; contact the INML Consortium (info@inml.nih.gov) for further information.

Insert Length: 845 Std Error: 0.00

Seq primer: reverse RT

High quality sequence step: 874.

Location/Qualifiers

FEATURES

SOURCE

1..464

/organism "Homo sapiens"

/db_xref "GDB:3794918"

/db_xref "taxon:9606"

/clone "IMAGE:145672"

/clone_lib "Soares fetal liver spleen INFLS"

/sex "male"

/dev_stage "20 week-post conception fetus"

/lab_host "DH108 (ampicillin resistant)"

/note "Organ: Liver and Spleen; Vector: pTZ19 (Pharmacia)

with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;

1st strand cDNA was primed with a Pac I - oligo(dT) primer

[5'- AACGGAGAAATTAATGAATCTTTTTTTTTTTTTT 4'].

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I

and Eco RI sites of the modified pTZ19 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 142 c 119 g 121 t 4 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Pref. No. 8; He-14;

Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4368 qqqagcttgaagcagaaatccttgaacccqaaagcagagcttgaagcagaa 4447

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

DB 107 GGAGAGGTGAGGCGAGCAATGGCTTGAACCGGAGCGAGCGAGGTTGACGTGAGCGAGA 48

QY 4448 tcaacacactgaactgaact 4469

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

DB 47 TCAACACACTGACTGCAAGCT 26

RESULT 8

AAZ25044

LOCUS

DEFINITION

AAZ24995

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

cDNA Library Preparation: David B. Krutzan, Ph.D.
cDNA Sequencing: Washington University Genome Systems Facility
clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.S.E. Consortium/INML at:
www.bio.lill.nyu.edu/Inmate/Inmate.html
Insert Length: 600 Std Error: 0.00
Seq primer: 41m153d, E1112g, Aroclor
High quality sequence step: 874.

Location/Qualifiers
1..464
/organism "Homo sapiens"
/db_xref "taxon:9606"
/clone "IMAGE:145672"
/clone_lib "NCI-CGAP_P42"
/sex "Male"
/dev_stage "45 weeks gest"

/lab_host "DH108"
/note "Vector: pTZ19 (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was ligated with a NotI - oligo(dT) primer [5'- AACGGAGAAATTAATGAATCTTTTTTTTTTTTTT 4'].

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with NotI and cloned into the NotI and Eco RI sites of the modified pTZ19 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 110 a 104 c 92 g 104 t 4 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Pref. No. 10; He-14;

Matches 81; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4385 aaagctgaagcagaaatccttgaacccqaaagcagagcttgaagcagaa 4447

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

DB 107 GGAGAGGTGAGGCGAGCAATGGCTTGAACCGGAGCGAGGTTGACGTGAGCGAGA 48

QY 4449 caacacacactgaactgaact 4469

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

DB 47 CAACACACTGACTGCAAGCT 27

RESULT 9

AAZ25044

LOCUS

DEFINITION

AAZ25044

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 464)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap/>

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgap@mail.nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquil, M.D., Michael Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krutzan, Ph.D.

cDNA Sequencing by: Genome Systems Facility

clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.S.E. Consortium/INML at:

47m153d, E1112g, Aroclor

High quality sequence step: 874.

Location/Qualifiers

1..464

/organism "Homo sapiens"

/db_xref "taxon:9606"

/clone "IMAGE:145672"

/clone_lib "NCI-CGAP_P42"

/sex "Male"

/dev_stage "45 weeks gest"

/lab_host "DH108"

/note "Vector: pTZ19 (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was ligated with a NotI - oligo(dT) primer [5'- AACGGAGAAATTAATGAATCTTTTTTTTTTTTTT 4'].

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with NotI and cloned into the NotI and Eco RI sites of the modified pTZ19 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 110 a 104 c 92 g 104 t 4 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Pref. No. 10; He-14;

Matches 81; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4385 aaagctgaagcagaaatccttgaacccqaaagcagagcttgaagcagaa 4447

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

DB 107 GGAGAGGTGAGGCGAGCAATGGCTTGAACCGGAGCGAGGTTGACGTGAGCGAGA 48

QY 4449 caacacacactgaactgaact 4469

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

DB 47 CAACACACTGACTGCAAGCT 27

RESULT 9

AAZ25044

LOCUS

DEFINITION

AAZ25044

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

—

```
QY 4375 atccacgactcggagagctgaagcaagaaatcgttgaaaccccaagcggagattg 4434
|||||
Db 285 ATCCGAGCTACTCGGAGGCTGAGCGAGGAATCGCTTGAACCCGGAGGAGGTTG 226
|||||
QY 4435 caatgaacccagatc 4449
|||||
Db 225 CAGTGAGCGGAGATC 211

RESULT 12
AW57414
LOCUS AW57414 526 bp mRNA linear EST 01-JUN-2000
DEFINITION EST369604 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AW57414
VERSION AW57414.1 GI:8147217
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 526)
AUTHORS Hendriks, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt,
J. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, I. J. and
Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johno@tigr.org
Plate: 107
Seq primer: Reverse.
FEATURES
Source
Location/Qualifiers
1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSKm"

BASE COUNT 150 a 108 c 150 g 117 t 1 others
ORIGIN

Query Match 1.0%; Score 75; DB 9; Length 526;
Best Local Similarity 100.0%; Pred. No. 8.3e-12;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4375 atccacgactcggagagctgaagcaagaaatcgttgaaaccccaagcggagattg 4434
|||||
Db 260 ATCCGAGCTACTCGGAGGCTGAGCGAGGAATCGCTTGAACCCGGAGGAGGTTG 319
|||||
QY 4435 caatgaacccagatc 4449
|||||
Db 220 CAGTGAGCGGAGATC 334

RESULT 13
AA527961
LOCUS AA527961 569 bp mRNA linear EST 05-AUG-1997
DEFINITION nh10e09.s1 NCI-CCAP_Px3 Homo sapiens cDNA clone IMAGE:953496
similar to contains Alu repetitive element; contains element MER29
MER29 repetitive element ;, mRNA sequence.
ACCESSION AA527961
VERSION AA527961.1 GI:2270030
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 569)
```

```
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap/
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ccapbs@mail.nih.gov
Tissue Procurement: W. Maistron Lincoln, M.D., Rodrigo Chuapin, M.D.,
Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David R. Kitzman, Ph.D.
cDNA Library Arrayed by: Genome Systems, Inc., Greg Lennon, Ph.D.
DNA Sequenced by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the T.M.A.C.E. Consortium/LINE at:
www-bio1.lnlni.gov/tbrp/image/line.html
Insert length: 564 Std Error: 9.00
Seq primer: -40mlx fwd. El from Amersham
High quality sequence stop: 587.
FEATURES
Location/Qualifiers
1..569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:953496"
/clone_lib="NCI-CCAP_Px3"
/sex="Male"
/dev_stage="45 years old"
/lab_host="Billerica"
/note="Vector: pAMbeta1 Site 1; N.T.S. Site 2; EcoRI; 1s"
strand cDNA was primed with oligo(dT)17 on 5' end of
UNAS-treated, total cellular RNA obtained from 5,000-10,
000 microdissected cells histologically determined to be
fully malignant prostate cancer cells. Double-stranded
cDNA was ligated to EcoRI adapters, 5 cycles of PCR
applied to the cDNA with an adapter-specific primer, and
the resulting PCR product subcloned into pAMP10 by the
BPE-cloning method (Life Technologies). Average insert
size is 600 bp. Note: Not directionally cloned. This
library was constructed by David Kitzman.
BASE COUNT 168 a 133 c 162 g 136 t
ORIGIN

Query Match 1.0%; Score 75; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 7.6e-12;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4375 atccacgactcggagagctgaagcaagaaatcgttgaaaccccaagcggagattg 4434
|||||
Db 351 ATCCGAGCTACTCGGAGGCTGAGCGAGGAATCGCTTGAACCCGGAGGAGGTTG 410
|||||
QY 4435 caatgaacccagatc 4449
|||||
Db 411 CAGTGAGCGGAGATC 425

RESULT 14
AF318360
LOCUS AF318360 2820 bp mRNA complete cds.
DEFINITION Homo sapiens pp7518 mRNA, complete cds.
ACCESSION AF318360
VERSION AF318360.1 GI:18027811
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2820)
AUTHORS Zhou, X.M., Zhang, P.P., Jiang, B.Q., Huang, Y., Qin, W.X., Zhao, X.T.,
Wan, D.F., and Gu, J.K.
TITLE Novel human cDNA clones with function of inhibiting cancer cell
growth.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2820)
AUTHORS Zhou, X.M., Zhang, P.P., Jiang, B.Q., Huang, Y., Qin, W.X., Zhao, X.T.,
```

Wan,D.F. and Gu,J.R.

TITLE Direct Submission
JOURNAL Submitted (02-NOV-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai 200032, P. R. China

FEATURES Location/Qualifiers

source 1. .2820 /organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1. .2820 /gene="pp7518"
CDS 1313..1621 /gene="pp7518"
/codon_start=1
/product="unknown"
/protein_id="AAL55867.1"
/db_xref="GI:18027812"
/translation="MYAGVGVGSPPHWWVGLTLPGLVRKSCYKTAKLLLVQRSPTFL
ATGSIFLEDHFSMEREGGAGFMEVFLPSSGISQTLLRSTQNLHSWCAQFTGGLLL

BASE COUNT 734 a 668 c 850 g 568 t

Query Match 1.0%; Score 75; DB 11; Length 2820;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4376 tccagctactcggagcctgagcagcagagaatcgctgaacccggaggcggaggttgc 4435
|||||
Db 955 TCCCAGCTACTCGGAGGCTGAGCAGGAGAATCGCTTGAACCCGGAGCGGAGGTTGC 1014
|||||
Qy 4436 agtgagccgagatca 4450
|||||
Db 1015 AGTGAGCCGAGATCA 1029
|||||

RESULT 15

BE061830 242 bp mRNA linear EST 09-JUN-2000
LOCUS RC1-BT0254-271199-014-g11 BT0254 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE061830
ACCESSION BE061830
VERSION BE061830.1 GI:8406480
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 242)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC1-BT0254-271
199-014-g11&t3-11-27&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 70.

FEATURES source

Location/Qualifiers
1. .242 /organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0254"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions." 47 t

BASE COUNT 68 a 54 c 73 g

Query Match 0.9%; Score 74; DB 9; Length 242;
Best Local Similarity 100.0%; Pred. No. 3.6e-11;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4376 tccagctactcggagcctgagcagcagagaatcgctgaacccggaggcggaggttgc 4435
|||||
Db 106 TCCCAGCTACTCGGAGGCTGAGCAGGAGAATCGCTTGAACCCGGAGCGGAGGTTGC 165
|||||
Qy 4436 agtgagccgagatc 4449
|||||
Db 166 AGTGAGCCGAGATC 179
|||||

Search completed: July 6, 2002, 15:24:07
Job time: 47166 sec

GenCore version 4.5
Copyright (c) 1994 - 2000 CompuGen Ltd.

QM nucleic acid sequence search, using SW model

Run on: July 6, 2002, 20:41:57 : Search time 1356.6 seconds
(without alignments)
8544.074 Million cell updates/sec

Title: US-09-761-466-5

Perfect score: 6751

Sequence: 1 caatttattnaqttrtat.....aatttaacccgagttttt 6751

Scoring table: OLIGO_NWC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :		N_geneseq_032802:*
1:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1960.DAT:*	
2:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1961.DAT:*	
3:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1962.DAT:*	
4:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1963.DAT:*	
5:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1964.DAT:*	
6:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1965.DAT:*	
7:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1966.DAT:*	
8:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1967.DAT:*	
9:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1968.DAT:*	
10:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1969.DAT:*	
11:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1970.DAT:*	
12:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1971.DAT:*	
13:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1972.DAT:*	
14:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1973.DAT:*	
15:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1974.DAT:*	
16:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1975.DAT:*	
17:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1976.DAT:*	
18:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1977.DAT:*	
19:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1978.DAT:*	
20:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1979.DAT:*	
21:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA1980.DAT:*	
22:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:*	
23:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:*	
24:	/SIDSI/accdata/hold-geneseq/geneseq-emb1/NA2002.DAT:*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB ID	Description
		Match	Length		
1	6751	100.0	6751	22 AAS09962	Genomic DNA #2 enc
2	640	9.5	972	22 AAH44455	Human NKX2-5/Csx n
3	640	9.5	972	22 AAH44224	Heart muscle cell
4	640	9.5	972	22 AAH49500	Human NKX2-5/Csx c
5	478	7.1	476	22 AAS09963	DNA encoding hcsx/
6	312	4.6	414	21 AA000441	Human MINT23 DNA c
7	194	2.9	1335	23 AAS78256	DNA encoding novel
8	49	0.7	729	23 AAS67506	DNA encoding novel
9	44	0.7	686	22 AA196691	Human neuroblastom

C 10	43	0.6	1525	28 AAX56674	Human secreted pro
C 11	43	0.6	1525	21 AAZ97114	Human secreted pro
C 12	43	0.6	1949	21 AAZ97334	Human secreted pro
C 13	42	0.6	138	21 AA77474	Human cancer assoc
C 14	42	0.6	208	21 AAZ14434	Human tumor express
C 15	42	0.6	236	21 AA17577	Human prostate can
C 16	42	0.6	331	21 AA17577	DNA encoding a tag
C 17	42	0.6	330	24 AHK6327	Human prostate can
C 18	42	0.6	377	22 AAK6327	Human immunoglob
C 19	42	0.6	410	22 AAG31174	Human polyubiquiti
C 20	42	0.6	724	21 AAG61266	Human secreted pro
C 21	42	0.6	758	22 AAH6674	Human polyubiquiti
C 22	42	0.6	767	22 AAH6675	Human polyubiquiti
C 23	42	0.6	818	22 AA134590	Human lung blastom
C 24	42	0.6	906	21 AA77512	Human cancer assoc
C 25	42	0.6	1036	20 AAV84494	Human secreted pro
C 26	42	0.6	1036	22 AHA83239	Human prostate can
C 27	42	0.6	1041	21 AAV15005	Human prostate can
C 28	42	0.6	1066	21 AAF21744	Human breast and o
C 29	42	0.6	1214	21 AAS96095	Human pancreatic c
C 30	42	0.6	1290	22 AAS26143	cDNA encoding for
C 31	42	0.6	1826	22 AAH44512	Human colon cancer
C 32	42	0.6	1854	22 AAS41114	cDNA encoding for
C 33	42	0.6	1907	22 AAB34845	Human colon cancer
C 34	42	0.6	1992	21 AAF23632	Human breast and o
C 35	42	0.6	3074	22 AAH45431	Human secreted pro
C 36	42	0.6	4419	22 AA159336	Human polyubiquiti
C 37	41	0.6	186	21 AAF22634	Human breast and o
C 38	41	0.6	253	21 AAH06697	Human colon cancer
C 39	41	0.6	264	21 AAH06697	Human colon cancer
C 40	41	0.6	285	21 AAH06697	Human colon cancer
C 41	41	0.6	294	22 AFA13431	Human colon cancer
C 42	41	0.6	308	19 AAV13431	Human secreted pro
C 43	41	0.6	368	19 AAV2365	Human R42453 cDNA
C 44	41	0.6	408	20 AAX06762	Novel nucleoside
C 45	41	0.6	352	22 AAS29344	cDNA encoding for

ALIGNMENTS

RESULT 1

AAS09962 standard; DNA: 6751 BP.

XX AAS09962;

XX AAS09962;

XX 24-JUL-2001 (first entry)

XX Becoming DNA #2 encoding human Csx/NKX2-5.

XX CSX/NKX2-5: cardiac enhancer for cardiac myoblast, myocyte, fibroblast, therapeutic heart tissue; sequence from human; 48.

XX Homo sapiens.

XX W020015106-A2.

XX 19-JUL-2001.

XX 16-JAN-2001; 2001W0-DS01511.

XX 14-JAN-2000; 2000DS-0176419.

XX (BETH-) BETH ISRAEL DEAN-NESS MEDICAL CENT.

XX Lee JW, Izumo S;

XX WPI; 2001-451809/48.

XX New cardiac specific cell enhancer elements, useful for specifically expressing gene in cardiac cell, as earlier marker of cardiomyocyte induction, e.g. for optimizing cardiomyocyte induction.

QY 1921 gcaaaacrtcaaggcccttctcccaaaadadadacacagaaagagagggaggttgaat 1980
DB 1921 gcaaaacrtcaaggcccttctcccaaaadadadacacagaaagagaggttgaat 1980
QY 1961 gcarqccrtccaaanaggggaaacaaagtgcctggtagctagggagcactccacaaacaa 2040
DB 1961 gcarqccrtccaaanaggggaaacaaagtgcctggtagctagggagcactccacaaacaa 2040
QY 2041 tatctccctgggaagctgaacttaccagctgaggtatctccatattatttcattatbaagaa 2100
DB 2041 tatctccctgggaagctgaacttaccagctgaggtatctccatattatttcattatbaagaa 2100
QY 2101 aagaagcattccctcggaaggtactagragtgactgagtcgccctggaggttgactctga 2160
DB 2101 aagaagcattccctcggaaggtactagragtgactgagtcgccctggaggttgactctga 2160
QY 2161 catagactactgaggaagacacaaactccaccccacactagccacactagccactggccc 2220
DB 2161 catagactactgaggaagacacaaactccaccccacactagccacactagccactggccc 2220
QY 2221 ttacttttcgttaattgcgccaaagaccccaagcctgacatgactgattgacctgtaga 2280
DB 2221 ttacttttcgttaattgcgccaaagaccccaagcctgacatgactgattgacctgtaga 2280
QY 2281 caaagcaaaccaaaccccgaattctccaaattttcgccctggtgtatccccaaagccc 2340
DB 2281 caaagcaaaccaaaccccgaattctccaaattttcgccctggtgtatccccaaagccc 2340
QY 2341 agccctgcttttaggttttttctattagattttccctatccctacccactttaat 2400
DB 2341 agccctgcttttaggttttttctattagattttccctatccctacccactttaat 2400
QY 2401 aataagccttccnaaaactaatttcccccacogcttccacccacttcttttttt 2460
DB 2401 aataagccttccnaaaactaatttcccccacogcttccacccacttcttttttt 2460
QY 2461 cccatgctggttgaggtactagaaatatttttcaaacacacacacacacacccctccc 2520
DB 2461 cccatgctggttgaggtactagaaatatttttcaaacacacacacacacacccctccc 2520
QY 2521 caaagccttaacttgcactgctctgtagatgttttcagggttacattagggagcaaaaq 2580
DB 2521 caaagccttaacttgcactgctctgtagatgttttcagggttacattagggagcaaaaq 2580
QY 2581 cagggtacaggggcaaaaagggagcccctccaaatggctgagcccttcaaaaagctga 2640
DB 2581 cagggtacaggggcaaaaagggagcccctccaaatggctgagcccttcaaaaagctga 2640
QY 2641 ggcagagnttt 2700
DB 2641 ggcagagnttt 2700
QY 2701 ttaqtdacacaaactgctcatgctcctgtcatcgaggccctggcccacatgacagcc 2760
DB 2701 ttaqtdacacaaactgctcatgctcctgtcatcgaggccctggcccacatgacagcc 2760
QY 2761 taagtccctctctctgactgctccagcctctcctgccccttgctcagcagctaccta 2820
DB 2761 taagtccctctctctgactgctccagcctctcctgccccttgctcagcagctaccta 2820
QY 2821 ctgcccacacacatccagagctgagcaggggtgagcagcagcagcagcagcagcagcag 2880
DB 2821 ctgcccacacacatccagagctgagcaggggtgagcagcagcagcagcagcagcagcag 2880
QY 2881 aaaaacttccagggagcctgagcaagcagccttctaccccacactggcctgtgagac 2940
DB 2881 aaaaacttccagggagcctgagcagcagccttctaccccacactggcctgtgagac 2940
QY 2941 taggcctgcccacacatgttccccagcctgctctcagcccacagccttctcagtaaaaa 3000
DB 2941 taggcctgcccacacatgttccccagcctgctctcagcccacagccttctcagtaaaaa 3000
QY 3001 catcctaaactggaagacagacagacagcagcagcagcagcagcagcagcagcagcag 3060

DB 3001 catcctaaactggaagacagacagcagcagcagcagcagcagcagcagcagcagcagcag 3060
QY 3061 cctagagagcagcctgagccttccctcctcctcctcctcctcctcctcctcctcctcctc 3120
DB 3061 cctagagagcagcctgagccttccctcctcctcctcctcctcctcctcctcctcctcctc 3120
QY 3121 cctagagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3180
DB 3121 cctagagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3180
QY 3181 ccttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3240
DB 3181 ccttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3240
QY 3241 ctacagagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3300
DB 3241 ctacagagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3300
QY 3301 acaagcctccttctccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3360
DB 3301 acaagcctccttctccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3360
QY 3361 aagacacacaccccttgaagcagagagagagagagagagagagagagagagagagagag 3420
DB 3361 aagacacacaccccttgaagcagagagagagagagagagagagagagagagagagagag 3420
QY 3421 gctctgattgtctgagcagcctcctcctcctcctcctcctcctcctcctcctcctcctc 3480
DB 3421 gctctgattgtctgagcagcctcctcctcctcctcctcctcctcctcctcctcctcctc 3480
QY 3481 cggattgaaagaaatccgaagacaaadacaaadacaaadacaaadacaaadacaaadacaa 3540
DB 3481 cggattgaaagaaatccgaagacaaadacaaadacaaadacaaadacaaadacaaadacaa 3540
QY 3541 cccagcagcttccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3600
DB 3541 cccagcagcttccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3600
QY 3601 aacaaacaaaaaaac 3660
DB 3601 aacaaacaaaaaaac 3660
QY 3661 cttttttaaatttttggadaaaadacacacacacacacacacacacacacacacacacac 3720
DB 3661 cttttttaaatttttggadaaaadacacacacacacacacacacacacacacacacacac 3720
QY 3721 ttt 3780
DB 3721 ttt 3780
QY 3781 ttt 3840
DB 3781 ttt 3840
QY 3841 gatatagaccccagagattttatt 3900
DB 3841 gatatagaccccagagattttatt 3900
QY 3901 caagcagaaaggttccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3960
DB 3901 caagcagaaaggttccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3960
QY 3961 aatagacag 4020
DB 3961 aatagacag 4020
QY 4021 gacccctgagacaaag 4080
DB 4021 gacccctgagacaaag 4080
QY 4081 atctccagaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 4140

Db 4081 attcccgaggccgacgagcgctgagggaagcgctggcggaaggccgctgcagccggg 4140
Qy 4141 agagaattcaagggtttgtgaggagcagagccctgggaacaaattcggcgggcgacggc 4200
Db 4141 agagaattcaagggtttgtgaggagcagagccctgggaacaaattcggcgggcgacggc 4200
Qy 4201 ggcagaaactgatcgctaccattcgaagaagccagcgaaggcaggtttccgagggccgctg 4260
Db 4201 ggcagaaactgatcgctaccattcgaagaagccagcgaaggcaggtttccgagggccgctg 4260
Qy 4261 cccaccgcgagctcttgagcaactgcgaacacctgcctgcggccaggtcgagctccga 4320
Db 4261 cccaccgcgagctcttgagcaactgcgaacacctgcctgcggccaggtcgagctccga 4320
Qy 4321 tcacaaaccacactccctggcctctgtttcttgattccttaattttgagataagacc 4380
Db 4321 tcacaaaccacactccctggcctctgtttcttgattccttaattttgagataagacc 4380
Qy 4381 gtccctagcagtgaagcctcgccctctgttatttaattctcaaaccaactagcccta 4440
Db 4381 gtccctagcagtgaagcctcgccctctgttatttaattctcaaaccaactagcccta 4440
Qy 4441 attcagttcacccagagcacaactcgttttatttttttttttttttttttttttt 4500
Db 4441 attcagttcacccagagcacaactcgttttatttttttttttttttttttttttt 4500
Qy 4501 ttttttttttttgagcctgaaatttaagtcaaccgtttgtctccctcacagagtggtg 4560
Db 4501 ttttttttttttgagcctgaaatttaagtcaaccgtttgtctccctcacagagtggtg 4560
Qy 4561 aactgccccagggcagagacctcccggtttgttttcagcgcccttgagcagcttgact 4620
Db 4561 aactgccccagggcagagacctcccggtttgttttcagcgcccttgagcagcttgact 4620
Qy 4621 ttttacaaatgctgagtgagacgtgtcggtggtccctcagtgcaactggcagagtgagcg 4680
Db 4621 ttttacaaatgctgagtgagacgtgtcggtggtccctcagtgcaactggcagagtgagcg 4680
Qy 4681 cagccagctggcgctccagcagacacagtgccctccacagaggtcccttaccattac 4740
Db 4681 cagccagctggcgctccagcagacacagtgccctccacagaggtcccttaccattac 4740
Qy 4741 tgtcgcccgcgctccgtaggtcaagcgctcttaccagcgctttctgctcttctgtt 4800
Db 4741 tgtcgcccgcgctccgtaggtcaagcgctcttaccagcgctttctgctcttctgtt 4800
Qy 4801 cccctcagagctgtgcgctgcagaagcggtggagctggagaagacagagcgcgaca 4860
Db 4801 cccctcagagctgtgcgctgcagaagcggtggagctggagaagacagagcgcgaca 4860
Qy 4861 acgcgagcgccccggcgcgacgcgcgaggaagcgcgcttcttcgcagggcg 4920
Db 4861 acgcgagcgccccggcgcgacgcgcgaggaagcgcgcttcttcgcagggcg 4920
Qy 4921 aggtctatgagctgagcgcgcttcaagcagcagcgaggtacctgtcgcccccaacg 4980
Db 4921 aggtctatgagctgagcgcgcttcaagcagcagcgaggtacctgtcgcccccaacg 4980
Qy 4981 accagctggcgagctgtgaaactcagtcaccagcaggtcaagatctggttccagaacc 5040
Db 4981 accagctggcgagctgtgaaactcagtcaccagcaggtcaagatctggttccagaacc 5040
Qy 5041 ggcgtacaagtgaagcgcgagcgcgagaccagactctggagctgtggggtgcccc 5100
Db 5041 ggcgtacaagtgaagcgcgagcgcgagaccagactctggagctgtggggtgcccc 5100
Qy 5101 cgcgcgcgcgcgcgcctgcgcgaagatcgcggtgcagtgctggtgcgcgagtggaagc 5160
Db 5101 cgcgcgcgcgcgcgcctgcgcgaagatcgcggtgcagtgctggtgcgcgagtggaagc 5160
Qy 5161 catgcttaggggactcggcgccctacgcgctgctacgcggtggcgctcaatccctacg 5220
Db 5161 catgcttaggggactcggcgccctacgcgctgctacgcggtggcgctcaatccctacg 5220

Qy 5221 gttataacccctaccgccctatccgggttacggcgcgcgccctgcagccctggctaca 5280
Db 5221 gttataacccctaccgccctatccgggttacggcgcgcgccctgcagccctggctaca 5280
Qy 5281 gctgactccgcttaccgcccgccctccccagcgagcgccactgcgcgcgcgcca 5340
Db 5281 gctgactccgcttaccgcccgccctccccagcgagcgccactgcgcgcgcgcca 5340
Qy 5341 acaacaacttcgtgaacttcggcgctcggggacttgatcggttcagagccccgggattc 5400
Db 5341 acaacaacttcgtgaacttcggcgctcggggacttgatcggttcagagccccgggattc 5400
Qy 5401 cgcagagcaactcgggagtgctcaacgctgcagtgtccgagcctggttaggaaggacc 5460
Db 5401 cgcagagcaactcgggagtgctcaacgctgcagtgtccgagcctggttaggaaggacc 5460
Qy 5461 cgcgtggcgccacctgaccgatccacactcaacagctccctgactctcgtggggagaag 5520
Db 5461 cgcgtggcgccacctgaccgatccacactcaacagctccctgactctcgtggggagaag 5520
Qy 5521 gggctcccaaatgacctgagtcacctggattttgatttcattcaactcctgcgagacctagg 5580
Db 5521 gggctcccaaatgacctgagtcacctggattttgatttcattcaactcctgcgagacctagg 5580
Qy 5581 aaacttttctgtcccaacgcgctttgtcttcgcgcacgggagagtttgtgcgcgcat 5640
Db 5581 aaacttttctgtcccaacgcgctttgtcttcgcgcacgggagagtttgtgcgcgcat 5640
Qy 5641 tgcagcgtgcaatgagtgatcctgcagcctggtcttagctgtccccccagagtgccc 5700
Db 5641 tgcagcgtgcaatgagtgatcctgcagcctggtcttagctgtccccccagagtgccc 5700
Qy 5701 tccgagagtcctatggcaccgccggttgaaactggagctggagctgcgcgagcgcc 5760
Db 5701 tccgagagtcctatggcaccgccggttgaaactggagctggagctgcgcgagcgcc 5760
Qy 5761 gagatctggcgccctatccgcgagccagggcgcgccggcgcccttgcctatctcgcc 5820
Db 5761 gagatctggcgccctatccgcgagccagggcgcgccggcgcccttgcctatctcgcc 5820
Qy 5821 gtcgcccccaacgcacccacccgatttatgttttaactattgctgtaagaaatgacg 5880
Db 5821 gtcgcccccaacgcacccacccgatttatgttttaactattgctgtaagaaatgacg 5880
Qy 5881 atccctctccattaaagagagtgctgtaaccccgcaactgtgcttcttcagcttgcg 5940
Db 5881 atccctctccattaaagagagtgctgtaaccccgcaactgtgcttcttcagcttgcg 5940
Qy 5941 cgcttcagaagcagagagtgctgccccgggactgctcagatctcagggcacaggc 6000
Db 5941 cgcttcagaagcagagagtgctgccccgggactgctcagatctcagggcacaggc 6000
Qy 6001 attcctcagacaaattgataacattgataactataaaactaaacctgctggaaccta 6060
Db 6001 attcctcagacaaattgataacattgataactataaaactaaacctgctggaaccta 6060
Qy 6061 ctggttcgctgcgggcaactttctgagattgctcatataaactccataaactccaaaaa 6120
Db 6061 ctggttcgctgcgggcaactttctgagattgctcatataaactccataaactccaaaaa 6120
Qy 6121 aaaaaaactccataaatttagaagctgagccccgggagaggtttaagacttacctgcgagc 6180
Db 6121 aaaaaaactccataaatttagaagctgagccccgggagaggtttaagacttacctgcgagc 6180
Qy 6181 aaatagccagtagctcgaactctggttaaatccagatgctcacttcagagaccgcc 6240
Db 6181 aaatagccagtagctcgaactctggttaaatccagatgctcacttcagagaccgcc 6240
Qy 6241 ttcctgtgctcccaagctccctccttgaaactcaatgtgtgccaggcaacggttccagg 6300
Db 6241 ttcctgtgctcccaagctccctccttgaaactcaatgtgtgccaggcaacggttccagg 6300

XX Homo sapiens.
OS WO200026401-A1.
PN 11-MAY-2000.
PD
PE 02-NOV-1999; 99WO-US25251.
PF
PG 03-NOV-1998; 98US-0106925.
PH 10-MAY-1999; 99US-0309175.
PI (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PJ Issa J, Baylin S, Toyota M;
PK WPI; 2000-365641/31.
PL Methylated CpG island amplification (MCA) used to determine DNA
PM Methylation status of CpG sites in a given locus, useful for diagnosis
PN of age-related disorders e.g. colon cancer and dementia -
PO Claim 47; Fig 6F; 100pp; English.
PP
PQ The patent discloses a method known as Methylated CpG island (CGI)
PR Amplification (MCA) to identify methylated CpG sites (cytosine residues
PS 5' of neighbouring guanine). The method uses a methylation-sensitive
PT restriction endonuclease SmaI, that cleaves unmethylated CpG sites and
PU its isoschizomer xmaI, that cleaves both methylated and unmethylated
PV CGIs. It involves amplification of closely spaced methylated SmaI sites.
PW This method is used to determine the DNA methylation status of CpG sites
PX in a given locus, in normal and neoplastic cells. It can be used for
PY diagnosis of age-related disorders, associated with methylation of CGIs,
PZ e.g. dementia, arteriosclerosis and diabetes mellitus, cell proliferative
QA disorders, e.g. cancer of the kidney, colon, breast, uterus, prostate and
QB lung, leukaemia, glioblastoma, astrocytoma and neuroblastoma. It is also
QC used to evaluate the cell response to a methylation modulating agent. It
QD can also be used to screen samples for the presence of hypermethylation
QE of specific genes. The present DNA sequence is the human MINT23
QF (Methylated IN Tumours) clone derived from colon cancer cell line, Caco2.
QG It is mapped to chromosome 5q34-35 and has 64% GC nucleotides. It
QH corresponds to the 3' noncoding region of the human homeobox gene, Csx.
QI
QJ Sequence 414 BP; 80 A; 133 C; 133 G; 68 T; 0 other;
QK
QL
QM Query Match 4.6%; Score 312; DB 21; Length 414;
QN Best Local Similarity 99.5%; Pred. No. 7.8e-84;
QO Matches 412; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QP 5391 cccgggattccgcagagcaactcggagtggtccacgtgcatggtatccgagcctggtag 5450
QD 414 CCCGGGATTCCGCAGAGCAACTCGGGAGTGTCCAGCTGTCATGGTATCCGAGCCTGGTAG 355
QE 5451 ggaaggaccgcgtggcgcgacctgacacgtatccacctcaacagctccctgaactcgcg 5510
QF 354 GCAGAGGACCCGCGTGGCGGACCCCTCACCGATCCACCTCAACAGCTCCCTGACTCTCG 295
QG 5511 tggggagaagggggtcccaacatgacctgagtcacctggattttggaattcactctgcg 5570
QH 294 GGGGGAGAGGGCGTCCCAACATGACCTGATGTCCTCCCTGGATTTCATTCACCTCTCG 235
QI 5571 gagacctaggaactttttctgtccacgcgcggtttgttcttcgcacgagagagttgtg 5630
QJ 234 GAGACCTAGGAACATTTTCTGTCCACGCGGTTTGTCTTCTTGCACGCGGAGTTTGTG 175
QK 5631 gcgcgattatgcagcgtgcaatgagtgatctcagcgtgtgtctttagctgtccccc 5690
QL 174 GCGGCGATTATGACGCTGCAATGAGTATCTCTGACGCTGTGTCTTAGCTGTCCCCCCC 115
QM 5691 aggagtgccctccgagagtgccatgggacccccgggttggaaactgggactgagctcgggca 5750
QO 114 AGGAGTGCCCTCCGAGAGTGCCATGGGCACCCCGCTTGGAACTGGGACTGGACTCGGCA 55

QY 5751 cgcaggccctgagatctggcgcgccattccgcgagccagggccgcggcccgccggg 5804
DB 54 CGCAGGCGCTGAGATCTGGCGGCCCATTTCCGCGAGCCAGGCGCGGCCCGGG 1
RESULT 7
AA578256
ID AAS78256 standard; cDNA; 1335 BP.
XX AC AAS78256;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #14060.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG14069.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 14060; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1335 BP; 250 A; 413 C; 410 G; 262 T; 0 other;
SQ

Query Match 2.9%; Score 194; DB 23; Length 1335;
Best Local Similarity 100.0%; Pred. No. 6.8e-49;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

or by determining the presence of mutations in the new genes. Specific uses are described for each of the 94 genes, based on which tissues they are most highly expressed in, and include developing products for the diagnosis of treatment of cancer, tumours, diseases of the immune system, and total deficiencies, blood disorders, allergies, developmental abnormalities, autoimmune diseases, inflammation, arthritis, Alzheimer's and cognitive disorders, schizophrenia, asthma, psoriasis, sepsis, skin disorders, arteriosclerosis, diabetes, cardiovascular disorders, kidney disorders, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for identifying their binding partners. The sequences shown in AA86334 to AA86585 represent fragments of the secreted proteins.

XX Sequence 1525 BP; 284 A; 502 C; 458 G; 276 T; 5 other;

Query Match: 0.68; Score 43; 18 21; Length 1525;
 Best Local Similarity: 100.0%; Pred. No. 0.00049;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2649 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCT 2691
 LK TT
 LK 1500 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCT 1458

RESULT: 12
 AA277047
 ID AA277044 standard: cDNA; 1949 BP.
 AC AA277044;
 XX 13-APR-2005 (first entry)
 XX Human secreted protein gene 16 cDNA clone HRFHC5, SEQ ID NO:27.
 XX Human; secreted protein; cancer; tumor; developmental abnormality;
 KW total deficiency; blood disorder; immune system disorder; inflammation;
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 KW arteriosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 KW digestive disorder; endocrine disorder; infection; AIDS; leukemia;
 KW therapy; ds.
 XX Homo sapiens.
 OS
 PN W09466041-A1.
 XX 23-DEC-1999.
 XX 15-JUN-1999; 9600-US13418.
 XX 16-JUN-1998; 980S-0089507.
 XX 16-JUN-1998; 980S-0089508.
 XX 16-JUN-1998; 980S-0089509.
 XX 16-JUN-1998; 980S-0089510.
 XX 22-JUN-1998; 980S-0090112.
 XX 22-JUN-1998; 980S-0090113.
 XX (HUMAN) HUMAN GENOME SCI INC.
 XX Robert SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA;
 P1 Sopper DR, Arner CA, Endress GA, Carter KC, Murotsuki M, Flor R;
 P1 Laffour DW, Olsen RS, Shi Y, Moore PA, Komatsu S G;
 XX WPI: 2500-104100/09.
 XX P-FSDB: AA94240.
 XX New isolated human genes and the secreted polypeptides they encode,
 P1 useful for diagnosis and treatment of e.g. cancers, neurological
 P1 disorders, immune diseases, inflammation or blood disorders
 XX Claim 1: Page 426-427; 586pp; English.

AA297019 to AA297137 represent 94 isolated human secreted protein genes.
 AA866215 to AA866333 are the secreted proteins encoded by the 94 human genes. This sequence represents a fragment of one of the human secreted proteins. The genes and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Also, their clinical applications can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 94 genes based on which tissues they are most highly expressed in, and include developing products for the diagnosis of treatment of cancer, tumours, diseases of the immune system, and total deficiencies, blood disorders, allergies, developmental abnormalities, autoimmune diseases, inflammation, arthritis, Alzheimer's and cognitive disorders, schizophrenia, asthma, psoriasis, sepsis, skin disorders, arteriosclerosis, diabetes, cardiovascular disorders, kidney disorders, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for identifying their binding partners. The sequences shown in AA86634 to AA86658 represent fragments of the secreted proteins.

XX Sequence 1949 BP; 345 A; 656 C; 521 G; 453 T; 5 other;

Query Match: 0.68; Score 43; 18 21; Length 1949;
 Best Local Similarity: 100.0%; Pred. No. 0.00049;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2649 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCT 2691
 LK TT
 LK 1947 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCT 1458

RESULT: 13
 AA2784347
 ID AA278434 standard: cDNA; 1949 BP.
 AC AA278434;
 XX 08-FEB-2001 (first entry)
 XX Human cancer associated gene sequence SE, US N 700.
 XX Human; cancer associated gene; cancer; antigen; biological cancer;
 KW diagnosis; cytostatic; proliferative tumour; immunomodulator;
 KW antidiabetic; antitumor; antihemorrhagic; antithrombotic; antiviral;
 KW antiinflammatory; antihypertensive; antidiabetic; cardiovascular;
 KW dermatological; neuroprotective; thrombolytic; coagulant; neoplastic;
 KW vasotropic; antipsoriasis; antileukemia; antitumor; inflammatory;
 KW immune disorder; haematopoietic cell; ds; cancer; autoimmune disorder;
 KW allergic reaction; graft versus host disease; viral infection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening; ds.
 XX Homo sapiens.
 OS
 XX W020055350-A1.
 XX 21-SEP-2000.
 XX 08-MAR-2000; 2005W-0505EP2.
 XX 12-MAR-1999; 94CS-0124270.
 XX (HUMAN) HUMAN GENOME SCI INC.
 XX Rosen CA, Kohn SM;
 XX WPI: 2500-58753/55.
 XX P-FSDB: AA844195.
 XX Novel isolated nucleic acids comprising sequences encoding peptides
 P1 useful for treating or diagnosing e.g. cancer

<hr/>					
AF135523	AF135523.1	GI:4914706			
VERSION					
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
REFERENCE	Toyota,M., Ho,C., Anub,N., Jair,K.W., Li,Q., Ohe-Toyota,M., Baylin,S.B. and Issa,J.P. Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification Cancer Res. 59 (10), 2307-2312 (1999)				
AUTHORS	Issa,J.-P.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-MAR-1999) Oncology, Johns Hopkins University, 424 N. Bond Street, Baltimore, MD 21231, USA				
MEDLINE	Location/Qualifiers				
REFERENCE	1..414 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="MINT23" /cell_line="Caco2" /tissue_type="colon cancer"				
FEATURES	misc_feature ..414 80 a 133 c 133 g 68 t /gene="Differentially methylated CpG island sequence"				
BASE COUNT	Query Match 4.4% Score 312; DB %; Length 414;				
ORIGIN	Best Local Similarity 99.5%; Pred. No. 2.3e-164; Indels 0; Gaps 0; Matches 412; Conservative 0; Mismatches 2;				
QY	5391	cgcgaattccgacgaactcgaggagtgtccacqctgcagatgatccgagcctgtag	5450		
DB	414	CCCGGGATTCGGCAGASCACCTCGGGAGTGTCCACGCTGCATGTATCGAGCCTGGTAG	455		
QY	5451	ggaaaggaccgcggtgpcgcacctgacacqatccccactcaacagctccctgactctcg	5510		
DB	354	GGAAGGGACC GGCGTGCGCGACCCCTCACCGATCCCAGATCCCACTCAACAGCTCCCTGACTCTCG	295		
QY	5511	tggagaagaaggggctcccacaatgacctgagtcctccttgattttgcattoactctcgcg	5570		
DB	294	GGGGGAGAAGSGGGCTCCAAATGACCTGAGTCGCCCTGGATTTCGATTCATCTCTCTCG	235		
QY	5571	qaqaactagaactttttctgtgccacgcgcgttttcttgcacacggagaaatttgtg	5630		
DB	234	CAGACCTAGGAACATTTTCTGTGTCACCGCGGTTGTCTCGCAGCGAGAGTTTGTG	175		
QY	5631	gcgcgaattatgcacgtgcgaatgaatgatcctgcagcctgatttctaactctcccc	5690		
DB	174	CGGGCAATATGCACGCTGCAATGAGTATCTCGACCTGGTGTCATTAGCTGCCCCC	115		
QY	5691	aggatgcctccgaagatccatggcacccccctgttgaaactggagctagctgggcca	5750		
DB	114	AGGAGTGGCCTCCGAGAGTCCATGGCACCCCGCGTTTGGAACTGGGAGCTCGGCA	55		
QY	5751	cgcagggcctgaatctatggcccccattccgcagacccagagcccgagcccgag	5804		
DB	54	CGCAGGCCCTGAGATCTGGCCGCTCATTCGCGAGCCAGGCGGGCGCCGGG	1		
<hr/>					
RESULT	7				
HS94ERR	HS4ERR				
LOCUS	H.sapiens Cpg island DNA genomic Msel fragment, clone 94e8, reverse				
DEFINITION	read cp94e8.rtl.a.				
ACCESSION	Z64004				
VERSION	Z64004.1 GI:10367382				
DESCRIPTION	Cpg island, genomic Msel fragment.				

XX WPI: 1994-341353/42.
XX
XX New regulatory regions of human erythropoietin gene - used for
PT studying and treating diseases and for prodn. of transgenic
PT animal models.
XX
XX Claim 1: Page 57-58; 81pp; English.
XX
XX AAQ79353 shows the nt. sequence of the entire 9.3 kb genomic clone
CC hEpsLH. This nucleic acid sequence includes a coding sequence, a 5'
CC flanking region contg. multiple regulatory elements and a 3'
CC flanking region contg. multiple regulatory elements. AAQ79353
CC consists of the last 1777 bases of AAQ79753. It corresp. to the non-
CC coding 3' flanking region of AAQ79753 and includes all the regulatory
CC elements contained therein. It extends from the 5'-most PstI site
CC 3' of the end of the coding sequence to a BamHI site. It comprises
CC 1777 bp exhibiting many stem-loop structures. It also contains TATA
CC boxes in forward and reverse orientation, and at least about 184
CC potential transcriptional regulatory elements. These elements
CC include TFIID, metal responsive elements, glucocorticoid
CC responsive elements (including GR/PP-PPV), NF-kappa-B, AP1, AP2,
CC Sp1, and lymphokine responsive consensus sequences, and many
CC others. This region also contains two nitrogen regulatory/oxygen
CC sensing sequences.
XX
XX Sequence 1777 BP: 400 A; 437 C; 453 G; 487 T; 0 other;

Query Match 0.9%; Score 74; DB 15; Length 1777;
Best Local Similarity 100.0%; Pred. No. 7.2e-19;
Matches 74; Conservation 0; Mismatches 0; Indels 0; Gaps 0;
QY 4376 tccagctactcggagctgaagcaggaggaatcgttgaaacccggagcgaagcttgc 4435
|||||
DB 1670 tccagctactcggagctgaagcaggaggaatcgttgaaacccggagcgaagcttgc 1611
QY 4436 agtgaacccagatc 4449
|||||
DB 1610 agtgagccgagatc 1597

RESULT 13
AAQ79353/c
ID AAQ79353 standard; DNA: 9272 BP.
XX
XX AC AAQ79353;

05-JUN-1995 (first entry)

DE Human genomic clone hEpsLH including erythropoietin (EPO) coding
DE sequence and 5' and 3' regulatory elements.
XX
XX Erythropoietin; erythropoiesis; red blood cell; regulatory element;
KW CAAT box; TATA box; ss.
XX
XX Homo sapiens.

XX Key Location/Qualifiers
FH CAAT_signal 875..881
FT /*tag= a
FT misc_signal 1754..1759
FT /*tag= b
FT TATA_signal 1954..1959
FT /*tag= c
FT TATA_signal 2006..2010
FT /*tag= d
FT TATA_signal 2304..2309
FT /*tag= e
XX
XX W09423570-A.
XX 27-OCT-1994.

XX 15-APR-1994: 94WO-US04141.
XX
XX 15-APR-1993: 93US-0046295.
PR 23-JUN-1993: 93US-0082850.
XX
XX (UUNY) UNIV NEW YORK STATE.
XX
XX Lee-Huang S;
XX
XX WPI: 1994-341353/42.
XX
XX New regulatory regions of human erythropoietin gene - used for
PT studying and treating diseases and for prodn. of transgenic
PT animal models.
XX
XX Disclosure: Page 51-55; 81pp; English.
XX
XX AAQ79353 shows the nt. sequence of the entire 9.3 kb genomic clone
CC hEpsLH. This nucleic acid sequence includes a coding sequence, a 5'
CC flanking region contg. multiple regulatory elements and a 3'
CC flanking region contg. multiple regulatory elements.
XX
XX Sequence 9272 BP: 2055 A; 2570 C; 2426 G; 2241 T; 0 other;
Query Match 0.9%; Score 74; DB 15; Length 9272;
Best Local Similarity 100.0%; Pred. No. 6.4e-19;
Matches 74; Conservation 0; Mismatches 0; Indels 0; Gaps 0;
QY 4376 tccagctactcggagctgaagcaggaggaatcgttgaaacccggagcgaagcttgc 4435
|||||
DB 9165 tccagctactcggagctgaagcaggaggaatcgttgaaacccggagcgaagcttgc 9160
QY 4436 agtgaacccagatc 4449
|||||
DB 9105 agtgagccgagatc 9092
RESULT 14
AAK87546
ID AAK87546 standard; DNA: 5211 BP.
XX
XX AC AAK87546;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42358.
DE
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
XX W0200157182-A2.
FN
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001: 2001WO-US01354.
XX
XX 31-JAN-2000: 2000US-0179065.
PR 04-FEB-2000: 2000US-0180628.
PR 24-FEB-2000: 2000US-0184664.
PR 02-MAR-2000: 2000US-0186350.
PR 16-MAR-2000: 2000US-0189874.
PR 17-MAR-2000: 2000US-0190076.
PR 18-APR-2000: 2000US-0198123.
PR 19-MAY-2000: 2000US-0205515.
PR 07-JUN-2000: 2000US-0209467.
PR 28-JUN-2000: 2000US-0214886.
PR 30-JUN-2000: 2000US-0215135.
PR 07-JUL-2000: 2000US-0216047.
PR 07-JUL-2000: 2000US-0216880.

SeqCore version 4.5
Copyright (c) 1995 - 2000 Compugen, Ltd.

CM nucleic - nucleic search, using sw model

Run on: July 6, 2002, 20:18:21 : Search time 249.41 Seconds
(without alignments)

6426.493 Million cell updates/sec

Title: US-09-761-466-5

Perfect score: 6751

Sequence: 1 caattctatttgaatttat.....auttaaacacccagatgatt 6751

Scoring table: 01430_NUC

Gapop 60.0 , Gapext 60.0

Searched: 48433 seqs, 122816752 residues

Word size : 6

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post processing: Listing first 45 summaries

Database : 1: /cnn2_6/ptodata/2/ina/VA_COMB.seq.*
2: /cnn2_6/ptodata/2/ina/BA_COMB.seq.*
3: /cnn2_6/ptodata/2/ina/HA_COMB.seq.*
4: /cnn2_6/ptodata/2/ina/HA_COMB.seq.*
5: /cnn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cnn2_6/ptodata/2/ina/BACKFILES.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	41	0.6	308	2	US-08-721-488-4
C 2	41	0.6	1134	3	US-09-248-345-29
C 3	41	0.6	1720	4	US-09-227-357-139
C 4	41	0.6	2209	1	US-08-514-014-1
C 5	41	0.6	2209	2	US-08-833-823-1
C 6	41	0.6	2280	3	US-08-813-150-1
C 7	41	0.6	2852	3	US-09-027-137-2
C 8	41	0.6	5962	6	5380255
C 9	41	0.6	5975	1	US-08-404-3648-1
C 10	41	0.6	5975	1	US-08-414-0838-1
C 11	41	0.6	5975	1	US-08-435-6753-1
C 12	41	0.6	5975	1	US-08-436-257A-4
C 13	41	0.6	5975	3	US-08-884-599-1
C 14	40	0.6	457	2	US-08-841-349-18
C 15	40	0.6	515	3	US-08-589-028-1
C 16	40	0.6	515	3	US-08-784-582-1
C 17	40	0.6	515	4	US-08-785-271-1
C 18	40	0.6	630	1	US-08-185-414E-1
C 19	40	0.6	790	4	US-09-461-474-14
C 20	40	0.6	806	4	US-08-955-629C-1
C 21	40	0.6	857	1	US-08-308-883-1
C 22	40	0.6	857	1	US-08-740-163-1
C 23	40	0.6	857	4	US-08-256-749-1
C 24	40	0.6	857	4	US-08-462-437-1
C 25	40	0.6	1023	1	US-08-252-966B-16
C 26	40	0.6	1056	1	US-08-157-101A-4
C 27	40	0.6	1307	2	US-08-961-022-17

C 28	40	0.6	1359	4	US-08-427-674-11
C 29	40	0.6	1359	4	US-08-427-674-11
C 30	40	0.6	1533	1	US-08-434-454A-1
C 31	40	0.6	1533	1	US-08-434-454A-1
C 32	40	0.6	1592	1	US-08-127-112A-1
C 33	40	0.6	1642	2	US-08-443-471-1
C 34	40	0.6	1642	2	US-08-443-471-1
C 35	40	0.6	1642	2	US-08-443-471-1
C 36	40	0.6	1683	4	US-08-432-463-11
C 37	40	0.6	1813	4	US-09-071-224-3
C 38	40	0.6	2144	4	US-08-923-454A-28
C 39	40	0.6	2700	3	US-09-341-99-1
C 40	40	0.6	4933	1	US-08-129-376-1
C 41	40	0.6	5933	3	US-08-461-731-1
C 42	40	0.6	5933	3	US-08-461-731-1
C 43	40	0.6	5933	3	US-08-461-731-1
C 44	40	0.6	4756	3	US-08-443-145-1
C 45	40	0.6	4780	3	US-08-443-145-1

ALIGNMENTS

RESULT 1

US-08-721-488-4/c
Sequence 4, Application US/08721488
Patent No. 5965388
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: LaValle, Edward
APPLICANT: Racie, Lisa
APPLICANT: Meberg, David
APPLICANT: Treary, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Rowan, Michael
TITLE OF INVENTION: SEQUENTIAL PEPTIDES AND POLYPEPTIDES
TITLE OF INVENTION: ENDO-LIN ITEM
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPILED: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatchIt, Release #1.1, Version #1.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721488
FILING DATE:
CLASSIFICATION No: 424
ALTERNATIVE/AGENT INFORMATION:
NAME: Rowan, Scott A. 27724
REGISTRATION NUMBER: 27724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8424
TELEFAX: (617) 876-5851
INFORMATION FOR SFO ID No: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 308 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLGY: linear
MOLECULE TYPE: cDNA
US-08-721-488-4

Query Match: 100%
Best Local Similarity: 100%
Seq No: 424
Pat. No.: 5965388

STRANDEDNESS: single
Topology: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...5700
OTHER INFORMATION:
US-09-761-466-5

Query Match 0.68; Score 41; DB 1; Length 5975;
Best Local Similarity 100.0%; Prod. No. 1e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

27 2649
Db 5969

RESULT 10
US-09-761-466-5-1/c
Sequence 1, Application US/09/814084B
Patent No. 5969241
GENERAL INFORMATION:
APPLICANT: Ellis, Steven Bradley
APPLICANT: Williams, Mark E.
APPLICANT: Harpold, Michael Miller
APPLICANT: Schwartz, Arnold
APPLICANT: Brenner, Robert
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McJain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/814,084B
FILING DATE: 28-SEPT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/604,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seligman, Stephanie L.
REGISTRATION NUMBER: 44,779
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
Topology: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...5700
OTHER INFORMATION:
US-09-761-466-5

Query Match 0.68; Score 41; DB 1; Length 5975;
Best Local Similarity 100.0%; Prod. No. 1e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2649
DB 5969

RESULT 11
US-09-761-466-5-1/c
Sequence 1, Application US/8456675B
Patent No. 571025
GENERAL INFORMATION:
APPLICANT: Ellis, Steven Bradley
APPLICANT: Williams, Mark E.
APPLICANT: Harpold, Michael Miller
APPLICANT: Schwartz, Arnold
APPLICANT: Brenner, Robert
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McJain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/8456,675B
FILING DATE: 05-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/814,084
FILING DATE: 28-SEP-1994
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
APPLICATION NUMBER: US 07/604,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seligman, Stephanie L.
REGISTRATION NUMBER: 44,779
REFERENCE/DOCKET NUMBER: 6362-53191
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
Topology: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...5700
OTHER INFORMATION:
US-09-761-466-5-1

Query Match 0.68; Score 41; DB 1; Length 5975;
Best Local Similarity 100.0%; Prod. No. 1e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2649

```

RESULT 13
US-08-884-599-1/c
; Sequence 1, Application US/08884599
; Patent No. 6013474
; GENERAL INFORMATION:
; APPLICANT: Ellis, Steven Bradley
; APPLICANT: Williams, Mark E.
; APPLICANT: Harpold, Michael Miller
; APPLICANT: Schwartz, Arnold
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS

```

```

: Patent No. 5935594
:
: GENERAL INFORMATION:
:
: APPLICANT: MISHRA, LOPA
:
: TITLE OF INVENTION: GENES CODING PROTEINS 1
:
: FILE REFERENCE: XX/PO47470US0
:
: CURRENT APPLICATION NUMBER: US/08/841,349B
:
: CURRENT FILING DATE: 1997-04-30
:

```

```

; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
;
; FILE REFERENCE: XX/PO4470US0
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
;

```


GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

Multiple - precise search, using sw model

Run on: July 6, 2002, 15:24:07 ; Search time 10026.1 seconds

(without alignments)
4088 102 Million cell updates/sec

2006/05/27 10:11:20: 00000000

Title: US-09-761-466-5

310.5 1.01104

Sequence: 1 caattcttattnaattctatat.....aqtllaaaccacqattttt 6751

Scoring table: 0,100_NUC

напор 60.0 , Capext 60.0)

Searched: 1-736207 seqs, 6748477542 residues

Word size : 3715 prom

Total number of bits satisfying chosen parameters:

Minimum PB seq length: 6

Maximum PB seq length: 2000000000

post-processing: listing first 45 summaries

Datatypes :

★
..
[...]
[...]
[...]

l: em est.b:★

2: em_esthlm: *

estimation: *

4: em_estm: *

```
em_estov: *
```

em_estpl:★

7: m_estro: *

4: em_hlc;★

rb estl:*

```

100:  ub_est[2] =

```

$$u_k = u_k^* \quad k \in \mathbb{N}.$$

*:str_qb :7

uri: SSD_Wi

14: cm qss iliv

ssb we : 5

16: em_qss_vrt

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Copy Match	Length	Dr	ID	Description
C 1	475	7.0	815	9	A1604745	AF065945 t1R2q11.x
C 2	455	6.9	713	9	AM65197	AF65197 t1q1q11.x
C 3	443	6.6	590	9	A1478839	t24e4.x
C 4	438	6.5	674	9	A1936496	A1936496 t24e4.x
C 5	425	6.3	557	9	AW091138	AW091138 t24q07.x
C 6	408	6.0	452	9	AU146160	AU146160 t24q07.x
C 7	399	6.0	406	9	AU146160	AU146160 t24q07.x
C 8	369	5.9	418	9	AU119953	AU119953 t24q07.x
C 9	363	5.8	806	10	A1337524	A1337523 t24q07.x
C 10	345	5.1	546	9	AF184417	AF184417 t24q07.x
C 11	316	4.7	628	10	AF725233	AF725233 t24q07.x
C 12	312	4.6	421	10	BF115950	BF115950 t24q07.x
C 13	309	4.6	545	9	A1245518	A1245518 t24q04.x
C 14	277	4.0	938	10	BF510466	BF510466 t1-H-B14
C 15	270	4.1	390	12	AO683260	AO683260 t24q07.x
C 16	270	4.0	938	10	BF151000	BF151000 t24q07.x
C 17	261	3.9	763	10	B1845590	B1845590 t24q08q1.x
C 18	205	3.0	1587	10	BM470688	BM470688 t24q08q1.x

C 18	137	2.0	948	10	BG11523	RS115223_602343612.7
C 19	127	1.9	1076	10	B1448434	B1448434_60304663.9
C 20	108	1.6	1166	10	B1448434	B1448434_60304663.9
C 21	60	0.9	73	10	W13452	W13452_4243111.1
C 22	53	0.8	597	9	AA716454	AA716454_42511084.5
C 23	44	0.7	681	16	BH4344_48	BH4344_48_507538.2
C 24	44	0.7	1437	10	BH421444	BH421444_48861340.7
C 25	43	0.6	165	9	A1877485	A1877485_48094112.4
C 26	43	0.6	400	10	BH522112	BH522112_2774555.8A
C 27	43	0.6	428	9	AA542247	AA542247_5034561.2
C 28	43	0.6	400	9	A1345445	A1345445_5034561.2
C 29	43	0.6	749	9	AV716454	AV716454_5034561.2
C 30	43	0.6	726	9	AV716454	AV716454_5034561.2
C 31	43	0.6	762	9	AV716454	AV716454_5034561.2
C 32	43	0.6	867	9	BH542248	BH542248_5034561.2
C 33	43	0.6	861	9	AV716454	AV716454_5034561.2
C 34	43	0.6	948	10	BH116485	BH116485_5034561.2
C 35	42	0.6	50	10	BH542248	BH542248_5034561.2
C 36	42	0.6	79	9	A1298440	A1298440_30030304.8
C 37	42	0.6	160	10	B28474	B28474_F0-2408_24.8
C 38	42	0.6	108	10	1949454	1949454_X742145.1
C 39	42	0.6	121	10	BH434423	BH434423_4817438.8
C 40	42	0.6	124	9	AA67374	AA67374_X742145.1
C 41	42	0.6	151	10	B1448434	B1448434_60304663.9
C 42	42	0.6	164	9	AA67374	AA67374_X742145.1
C 43	42	0.6	168	9	AA142816	AA142816_80150111.8
C 44	42	0.6	174	9	AA473654	AA473654_80759411.1
C 45	42	0.6	181	9	AA194245	AA194245_80759411.1

ACKNOWLEDGMENTS

```

FEATURES             1
A1609745 /c        EST 12-MAY 1998
LOCUS              Homo sapiens cDNA clone IMAGE106945
DEFINITION         similar to SW-HK25_HUMAN pI252 BOMBIKX HOME BOX PROTEIN NKX-2.5 (containing
                    A1n repetitive element), mRNA sequence.
A1609745            Homo sapiens
A1609745.1         GI:5418942
EST                Human.
ORGANISM           Homo sapiens
Eukaryotic; Metazoa; Chordata; Craniota; Vertebrata; Epiosteiomi;
Mammalia; Eutheria; Primates; Artiodactyla; Homidae; Homo.
(cases 1 to 67)
N1/NINDS-CGAP http://www.ncbi.nlm.nih.gov/cgap/.
National Cancer Institute / NCI Data. Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/HGAP). Tumor Gene Index
Unpublished (1998)
Contact: Robert Strauchen, Ph.D.
Email: chaps1@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Keith P. Weststead, M.D.,
Ph.D.
CGMA Library Preparation: M. Bentz-Sorensen, Ph.D., M. Fatima
Benadid, Ph.D.
CGMA Library Arrayed by: Fred Leach, Ph.D.
cDNA Sequenced by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distributed information can be
found through the I.M.A.G.E. Consortium/http://immi.dcc.tohoku.ac.jp/bio/lib/arraylibrp/home/image.html
Insert Length: 1594 bp (first run)
Sequencer: 400p base pair
High quality sequence start: 45'
polyA No.          Location/Qualifiers
1..3815
/atomize "Homo sapiens"
/B_Xref "taxon:9606"
/clone "IMAGE106945"
/genome lib="NCI-CGAP-Plex"
```

/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(df) primer [5' TGTTACCAATCTGAAGTGGGCGGCGCATATCTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 157 a 245 c 275 g 131 t 7 others
ORIGIN

Query Match 7.0%; Score 475; DB 9; Length 815;
Best Local Similarity 99.7%; Pred. No. 7e-84;
Matches 575; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 5339 caacacaactctgaacttcggtcgggacttgaatcggttcagagcccgagat 5398
|||||
Db 577 CAACAACAACCTCTGGAACCTTCGGCGTTCGAGTTCGAATCGGTTTCAGACNCCGGGAT 518
Qy 5399 tccgcagagcaactcggagtgctccacgctgcatggtattccgagcctggtagggaaggga 5458
|||||
Db 517 TCCGCAGAGCAACTCGGAGGTGTCACGCTGCATGGTATCCGAGCCTGGTAGGGAAGGGA 458
Qy 5459 ccgcgtggtcgagccctgacgcatccacactcaacagctccctgactctcgtggggaga 5518
|||||
Db 457 CCCCGGTGGCGGACCTGACCGGATCCACCTCAACAGCTCCCTGACTCTCGGGGGGAGA 398
Qy 5519 aggggtcccaacatgatccctgagtccttggtattgcatcactcctcgcgagacctga 5578
|||||
Db 397 AGGGGCTCCCAACATGACCTGAGTCCCTCGGATTTTGCAATTCATCTCTCGCGAGACCTA 338
Qy 5579 ggaacttttctgtccacgcgcgtttgttcttgcacggagagatttggggggcgat 5638
Db 337 GGAACCTTTTGTGTCGCCACGGCGTGTGTTCTTGGCAGCGGAGAGTTTGTGGCGGCGAT 278
Qy 5639 tatcagcgtcaatgagtatcctgacgctggtgtcttagctgtcccccagagatgc 5698
|||||
Db 277 TATCAGCGTGCATATGATGATCTGACGCTGGTGCTTAGCTGTCTCCCGCAGAGTGC 218
Qy 5699 cctccgagagtcctatgggcaaccccggttggaactgggactgagctcgggcacgcagggc 5758
Db 217 CCTCCGAGAGTCCATGGCACCCCGGTGGAACTGGGACTGAGCTCGGCGCACGAGGC 158
Qy 5759 ctgagatctggcgccattccgcgagccagggcggcgcccggttgcttgcctatctcg 5818
Db 157 CTGAGATCTGGCGCCCATTTCCGCGAGCCAGGGCGCGCGGCGCTTTGCTATCTCG 98
Qy 5819 ccgtcgccgcgccacgcaccacccgtatttatgttttacctattgtctgaagaaatga 5878
Db 97 CCGTCGCCGCCGACGACCCACCCGCTATTTATGTTTATCTATTGCTGTGAAGAAATGA 38
Qy 5879 cgatcccttccattaaagagagtgogttgaccccg 5915
Db 37 CGATCCCGCTTCCCATTAAGAGAGTGCCTTGACCCCG 1

RESULT 2
AW665197/c
LOCUS
DEFINITION h301g01.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2980560 3' similar to SW:HK25_HUMAN P52952 HOMBOBOX PROTEIN
NX-2.5 ; contains Alu repetitive element;; mRNA sequence.
ACCESSION AW665197
VERSION AW665197.1 GI:7457743
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 713)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 466.
Location/Qualifiers
1. 713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2980560"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I M A G E clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 145 a 205 c 236 g 125 t 2 others
ORIGIN

Query Match 6.9%; Score 466; DB 9; Length 713;
Best Local Similarity 99.6%; Pred. No. 4.6e-82;
Matches 566; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 5326 ccactgcgcgcgcacaacaacttcgtgaacttcggcgctcgggacttgaaatcggttc 5385
Db 584 CCAGTCCCGCGCCCAACAACAACACTTCGTCACTTCGGGCTCGGGACTTGAATCGGGTTC 525
Qy 5386 agagccccggagattccgcagagcaactcgggagtgatccacgtgcatgataccagact 5445
Db 524 AGAGCCCGGGATTCGCGAGAGCAACTCGGGAGTGTCCACGTGCATGATGATCCGAGCCT 465
Qy 5446 ggtagggaagggaccgcgctggtgcgcacctgacgagtcacccctcaacagctccctgac 5505
Db 464 GGTAGGGAAGGAGACCCCGTGGCGGACCCCTGACCGATCCACCTCAACAGCTCCCTGAC 405
Qy 5506 tctcgtgggagaagggtcctcccaacatgacctgagtcctcctgagtttgcaattcaactc 5565
Db 404 TCTCGGGGGAGAGAGGGGTCTCCAAACATGACCTGAGTCCCTGGATTTTGATTCATCTC 345
Qy 5566 ctggggagacctagaaactttttctgccacgcgctttgtcttcgcacagagagat 5625
Db 344 CTGCGGAGACCTTAGAACTTTTCTGTCCACGCGGTTTGTCTTTCGCGACGCGGAGAGT 285
Qy 5626 ttgtggtcgcatattatgacgctgcaatgagtgcctcctcagcctggtgtcttagctgtc 5685
Db 284 TTGTGCGCGGATTTATGACGCTGCAATGAGTGTCTCTCAGCCTGTCTCTTAGCTGTC 225
Qy 5686 cccccagagtgccctccgagagtcctatgggcccccccggttggaaactgggactgagctc 5745
Db 224 CCCCCAGGAGTGCCTCCGAGAGTCCATGGGCACCCCGCGTTTGGAACTGGGACTTGAGCTC 165
Qy 5746 gggcagcagggcctgagatctgcccgccttcacgagccagagccgagccgagccgagcc 5805
Db 164 GGGCAGCAGGGCCCTGAGATCTGGCGGCCCATTTCCGCGATCCAGGCCCGGCCCGCCGCG 105
Qy 5806 ctttctctctcgtcgtgcgcgcgcacacgaccccccatttatgttttttacctattg 5865
Db 104 CTTTGTCTATCTCGCGTGCSCCGGCCACCGACCCACCCCGCTATTATGTTTATACCTATTG 45

[illegible]

LOCUS AI337523 418 bp mRNA linear EST 13-FEB-1999
DEFINITION qq02a02.x1 Soares fetal_lung_NbHL19W Homo sapiens cDNA clone
IMAGE:1931306 3', mRNA sequence.
ACCESSION AI337523
VERSION AI337523.1 GI:4074450
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 418)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 606 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 366.
Location/Qualifiers
FEATURES
source
1..418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1931306"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: p7T73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACATCTGAAGTGGAGCGCGCAATTTTCTTTTCTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBHL19W."
BASE COUNT 96 a 120 c 131 g 71 t
ORIGIN
Query Match 5.9%; Score 399; DB 9; Length 418;
Best Local Similarity 100.0%; Pred. No. 8.9e-69;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5512 ggggagaagggtcccaacatgacctgagtcacctgatttgcatcactcctgcgg 5571
DB 400 GGGGAGAAGGGGCTCCCAACATGACCTGATCCCTCGATTTCATTCACTCTGGGG 341
QY 5572 agacctaggaactttttgtcccaacgctgtttgtcttcgcacgggagatttgtgg 5631
DB 340 AGACCTAGGAACCTTTTCTGTCCACGCGCTTTTCTTTCGCGACGGGAGATTGTGG 281
QY 5632 cggcgattatcagcgtgcaatgagtatcctgcagctgctgtcttagctgtccccc 5691
DB 280 CGGCGATTATCAGCGTGCAATGATGATCTGTGACCTGTGTCTTAGCTGTCCCCCA 221
QY 5692 ggagtgcctccgagatcccatgggacccccgggttgaaactgggaactggagctcgggac 5751
DB 220 GGAGTGCCCTCGGAGATCCATGGCACCCCCGGTTGGAACCTGGGACTGAGCTGGGCAC 161
QY 5752 gcagggocctgagatgtggcccccattccgcagaccagggccggcgccgctttgc 5811
DB 160 GCAGGGCCTGAGATGTGGCCGCCCATTCGCGAGACAGGCGCGCGCGCTTTTGC 101
QY 5812 tatctgcgcgtcgcccgccagccaccacccgtatttatgttttacctatttctgttaa 5871
DB 100 TATCTGCCGTCGCCGCCACGACACCCACCCGATTTATGTTTTTACCTATTGCTGTAA 41

QY 5872 gaaatgacgtcccccttccattaaagagagtgcttga 5910
DB 40 GAAATGACGATCCCTTCCCATTAAGAGAGTGGCTGA 2
RESULT 9
BI834117 806 bp mRNA linear EST 04-OCT-2001
LOCUS 603085284F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5224365 5',
DEFINITION mRNA sequence.
ACCESSION BI834117
VERSION BI834117.1 GI:15945667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 806)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11564 row: g column: 22
High quality sequence stop: 736.
Location/Qualifiers
FEATURES
source
1..806
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5224365"
/clone_lib="NIH_MGC_120"
/lab_host="DH10B"
/note="Organ: pooled pancreas and spleen; Vector:
pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 025. Note: this is a NIH_MGC Library."
BASE COUNT 135 a 290 c 272 g 109 t
ORIGIN
Query Match 5.8%; Score 393; DB 10; Length 806;
Best Local Similarity 100.0%; Pred. No. 6e-68;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2895 gccgtgggcagcgcttctgcgcacacctggcgctgtgagactggcgctgcacca 2954
DB 53 GCCGTGGGCAGCGCGCTTCTTCCGCCACCTGGCGCTGTGAGACTGGCGTCCACCA 112
QY 2955 tgttccccagcctgtctctcagcccaacccttctcaglcacaagacatcctaaacctgg 3014
DB 113 TGTTCCTCCAGCCCTGTCTCAGCGCCACGCCCTTCTCAGTCAAGAGACATCTTAAGCTGG 172
QY 3015 aacacagcagcgagcctgctgcgcgagagctctctgcgcacctggagcgaccc 3074
DB 173 AACAGCAGCAGCGCAGCCTGGCTGCCGCCGAGAGCTCTCTGCCGCCCTTGGAGCGACCC 232
QY 3075 tggegcctccctcctgcacgtggcgcccttcaagccagagcgctacgtcggtggccggag 3134
DB 233 TGGCGCCCTCCTCTGCTGCTGCTGGCGCCTTCAAGCCAGAGGCGCTACGCTGGGCCGAGG 292

FEATURES	PhyA Yes	Location/Qualities	Accession	AC083250					
Source	1: 555	Jordanism-"Homo sapiens"	AC083250.1	GI:56414757					
		/db_xref="taxon:9606"	GSS						
		/clone="IMAGE:3084892"	ORGANISM						
		/lab_host="NCL_CGAP_Sub8"	human						
		/note="Vector: pT7Rb-Pac (Pharmacia)"							
		/polylinker: Site1: Not 1; Site2: Eco RI; NCL_CGAP_Sub8							
		is a subtracted library derived from NCL_CGAP_Sub5. The							
		single-stranded DNA preparation of NCL_CGAP_Sub5 was used							
		as a tracer in a subtractive hybridization with a driver							
		comprising a pool of clones from NCL_CGAP_Sub5 (IMAGE							
		clone ids 2742844-2747415, 3068040-3069191); 25% of the							
		(IMAGE clone ids 2724592-2729926; 25% of the driver							
		population). NCL_CGAP_Sub6 (pool A1F-A1J, IMAGE ids							
		2728969-2733190; 25% of the driver population), and							
		NCL_CGAP_Sub7 (IMAGE ids 3059192-3072246, 3081844-3084550							
		; 25% of the driver population). Subtraction was							
		performed as previously described [Bonaldi, Leizoe, &							
		Scares (1996): Normalization and Subtraction: Two							
		Approaches To Facilitate Gene Discovery. Genome Research,							
		6, 791-806.							
BASE COUNT	115 a	161 c	164 t	110 T					
ORIGIN				1 others					
Query Match	4.6%	Score 409;	DB 10;	Length 555;					
Best Local Similarity	99.8%	Prod. No. 1.9e-51;							
Matches	424;	Conservative	0;	Mismatches	0;	Indels	1;	Gaps	1;
QY	5381	atctcaagaccccgattccggaaggaactccgaatctccagctatccagctatatacc	5443						
Db	545	ggttcacagggccgggattccgagagAACGAGACGCGGAGTGTCACGCTGATGGTATCG	486						
QY	5441	agcctatgaaagggacccgattgagagacccctaaacgatccagctcaacatccc	5500						
Db	485	AGCGTGGTAGGAGGGACCGCGCTGGCGGACCTGACCGATCCCACTCAACAGCTCC	426						
QY	5501	ctgactctctgaggaaggaagactccaaatgacctgacctgagctccctgatttgcatt	5560						
Db	425	CTGACTCTCTGCGGAGAGGGGCTCGCAACATGACCTTGGTGGCTGGATTTGGATT	366						
QY	5561	caactcaggaagacctagggaattttctgtccagacacgtttctcttctgacacag	5620						
Db	365	CACCTCTCTGGAGACCTAGGAA-TTTTCTGTGGCAGCGCGTTTCTGTGGCAGCGG	307						
QY	5621	aaatttatagcgcattatagacatcgaatgattgatacctgacctgattctag	5680						
Db	306	AGAGTTTGTGGTGGCCATTATGGAGGTGGCAATGAGTGATCTCCAGCTTGTGTCTTAG	247						
QY	5681	ctgtcccccaggaatgcctcccgaagtccatggacaccccgcgttgaacttgaactg	5740						
Db	246	CTGTCCGCCCTAGGAGCTGCCCTCCGAGAGTCCATGGCAGCCCGGTTGGAACTGGGACTG	187						
QY	5741	agctcagacacagagccctgaatctgacccacatctcgacagcagccagagccgc	5800						
Db	186	AGTTCGAGGACGACAGGGGCTGAGATCTGGTGGTGGATTCGATTCGGAGGCAAGGCGGAGGCGCC	127						
QY	5801	caagcctttt	5810						
Db	126	GGGAGCTTTT	117						
RESULT	14								
AC083250									
LOCUS	BS.5469.B1.B09.T7A.RP1-11	490 bp	DNA	linear	GSS	99-NV-1999			
DEFINITION	genomic clone Plato-5237 Col 17 Row-D, DNA sequence.								
KEYWORDS	Source: SSS								
ORGANISM	human								
REFERENCE	1 (bases 1 to 490)								
AUTHORS	Mahairas G.S., Wallace G.S., Smith K., Swartzell S., Holman J., Koller A., Shaker K., Friedman, Yehia I., Chao S., Adams M.F. and Hood L.								
TITLE	Sequence-tagged cDNAs: A sequence approach to mapping and								
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17): 9744-9744 (1999)								
MEDLINE	99360589								
COMMENT	Contact: Mahairas G.S., Wallace G.S., Smith K., Swartzell S., Holman J., Koller A., Shaker K., Friedman, Yehia I., Chao S., Adams M.F. and Hood L. High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98199, USA Tel: (206) 616 6634 Fax: (206) 616 6847 Email: jwallace@u.washington.edu. Clones are derived from the Human BAC Library PPL-11. BAC library availability: please contact the PPL-11 BAC library (pilot@u.washington.edu). Clones may be purchased from BAC200 Resources (http://BAC200Resources.org/) or from Rosetta Biosciences (415) 750-5000. BAC and Web Sites: http://www.hsc.washington.edu Plate: 9237 Row: L Column: 17								
Seq primer: 17									
Class: BAC ends									
High quality sequence status: 39									
Location/Qualities									
1: 490									
Jordanism "Homo sapiens"									
/db_xref="taxon:9606"									
/clone="Plato-5237 Col 17 Row-D"									
/sex="male"									
/note="Vector: pBAC1.9; Site 1: EcoRI; Site 2: PstRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and PstRI. NotI. Size selected DNA was cloned into the pBAC1.9 vector at EcoRI sites." 1 others									
BASE COUNT	84 a	106 c	86 t	115 T					
ORIGIN									
Query Match	4.3%	Score 273;	DB 12;	Length 490;					
Best Local Similarity	100.0%	Prod. No. 4.0e-43;							
Matches	277;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	4222	attcgaagaagcagcaagaaatctccagagagccagagagagagagagagagagag	4281						
Db	97	ATTCCGAGGAGGACACAAACAGCTTCGAGAGTGGCTGGAGTGGAGCTTCTTCAAT	156						
QY	4282	attgcacacacccctcttgcacacacacacacacacacacacacacacacacacac	4341						
Db	157	ACTCCGCAAAACCTGCTCCGCAATAGGCTGSACTTCGCAATAGGCAATAGGCTG	216						
QY	4422	actttctgtttctgatttttaattttttttttttttttttttttttttttttttttt	4481						
Db	217	actttctgtttctgatttttaattttttttttttttttttttttttttttttttttt	279						
QY	4402	actttctgtttctgatttttaattttttttttttttttttttttttttttttttttt	4461						
Db	277	actttctgtttctgatttttaattttttttttttttttttttttttttttttttttt	4461						
QY	4462	actttctgtttctgattt	4499						
Db	437	CANCTGGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	475						


```

/note="assembly_fragment:00204
fragment_chain:1"
misc_feature
47616..52829
/note="assembly_fragment:02002
fragment_chain:1"
misc_feature
52930..90493
/note="assembly_fragment:00959
fragment_chain:1"
misc_feature
90594..109805
/note="assembly_fragment:01522
fragment_chain:1"
misc_feature
109906..136657
/note="assembly_fragment:01742
fragment_chain:1"
BASE COUNT 33294 a 32671 c 32974 g 37117 t 601 others
ORIGIN
Query Match. 74.7%; Score 357; DB 2; Length 136657;
Best Local Similarity 99.8%; Pred. No. 1.3e-195;
Matches 477; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 aaaaaatcattaccgcgattcacaagaacatagagaatgaacagtcactatctttt 60
Db 46483 AGCAAAATCATTTACCGGATTCACAAAGAGCATAGAGGTGAACAGTCACTCTTGT 46424
QY 61 caaataggagagcttttttctctccctttttgtaacacctgaccacagagactacac 120
Db 46423 CAATAGGAGAGGTTTTTTCCTTCCTTTTGTAAACACCTGACCACAGGACTACAG 46364
QY 121 ttctagaaccccttaccgaaataaagaataaatactcttccacttgatttgcac 180
Db 46363 TTCAGAGCCGCTTACCCGAAATAGGAATAAATGCTTGCCACGCTGATTTGCAAG 46304
QY 181 gcaatcactaatttttttcttccagagctctcacaacacacacacacacacacac 240
Db 46303 GCAATGCTAATTTTTTCTTCCTCCAGAGCTCTC-AAAAAAAAAAAAAAAAACCTTAC 46245
QY 241 taaaaaagagatcccgatatacctcgaataatcccatcattaaagcgaataatttcac 300
Db 46244 TAAACAGGATGCTGGATGAGCTCGATGAGCTGCGCCCATTAACGGGTAAATTTACGGC 46185
QY 301 atcactcactaactatctttcaaaactttatcgcagagcgcctgacacagactacatt 360
Db 46184 GTCCGCTCACACTAATCTTCAAACTGCTATCGCAGCGCGCTGGCAGCAGATTCACCT 46125
QY 361 aacagactcccaaacctcgttccgaactcttttcaagacacatttaattgaatcgg 420
Db 46124 AACAGGCTCCGAGGACCGCTGTCGAGGCTTTTTCAGCGAGACATTTAATTAATGATCGG 46065
QY 421 atatagctcttttgcacacactcaccacctcgcagatagagatcctctccacagacac 478
Db 46064 ATGTGGCTCGTTTGGCAGACGTCACGCGCTCGCGGATAGGCACTCTCCACAGACAC 46007
RESULT 3
AL669821 AL669821 141702 bp DNA linear HTG 31-JAN-2002
LOCUS Homo sapiens chromosome 5 clone XXbac-116A1, *** SEQUENCING IN
DEFINITION PROGRESS ***, 4 unordered pieces.
ACCESSION AL669821
VERSION AL669821.4 GI:18157151
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FUILLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
REFERENCE Mashreghi-Mohammadi, M.
AUTHORS Direct Submission
TITLE Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

```

COMMENT

```

On Jan 15, 2002 this sequence version replaced at:18152648.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bP116A1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 106752; 100% of reads
Chemistry: Dye-terminator 334 bp; 106% of reads
Consensus quality: 144176 bases at least 44%
Consensus quality: 144232 bases at least 60%
Consensus quality: 144299 bases at least 92%
Insert size: 144402; sum-of-contents
Quality coverage: 167909; 14.3% error; average-if
quality coverage: 16728 in 620 bases; sum-of-contents quality
coverage: 9.55% in 620 bases; average-if
-----

```

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 75929 76028: gap of 100 bp

* 76029 116346: contig of 400 bp in length

* 116347 116446: gap of 100 bp

* 116447 131258: contig of 14812 bp in length

* 131259 131358: gap of 180 bp

* 131359 144702: contig of 13444 bp in length.

FEATURES

source

1..144702

organism="Homo sapiens"

db_xref="taxon:9606"

chromosome="5"

clone="XXbac-116A1"

clone_jit="CH01-5-1"

1..75928

/note="assembly_fragment:00204"

fragment_chain:1

clone_end:17

vector_side:left"

76029..116346

/note="assembly_fragment:00959"

fragment_chain:1

116447..131258

/note="assembly_fragment:00954"

fragment_chain:1

131359..144702

/note="assembly_fragment:01454"

fragment_chain:1

clone_end:516

vector_side:right"

BASE COUNT 41517 a 33170 c 32886 g 36829 t 601 others

ORIGIN

Query Match. 74.7%; Score 357; DB 2; Length 144702;

Best Local Similarity 99.8%; Pred. No. 1.3e-195;

Matches 477; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 aagaatcattaccgcgattcacaagaacatagagaatgaacagtcactatctttt 60

Db 11094 AGCAAAATCATTTACCGGATTCACAAAGAGCATAGAGGTGAACAGTCACTCTTGT 11153

QY 61 caaataggagagcttttttctctccctttttgtaacacctgaccacagagactacac 120

Db 11154 CAATAGGAGAGGTTTTTTCCTTCCTTTTGTAAACACCTGACCACAGGACTACAG 11213

QY 121 ttctaggaagcccccttaccgaaataaggaataataatccttgccaccttgatttgcaag 180
|||||
Db 11214 TTCTAGGAAGCCCCCTTACCGAAATAGGAAATAAATCCTTGCCACCTTGATTGCAAG 11273
|||||
QY 181 ggcaatgctaatttttttcttctccagagctctcaaaaaaaaaaaaaaaaaaccccttac 240
|||||
Db 11274 GGCAATGCTAATTTTTTCTCTCCAGAGCTCTC-AAAAAAAAAAAAAAAAACCTTAC 11332
|||||
QY 241 taaaacagggatcccgatgtagcctcgtatgtcccccattaaacggtaatatatttcaggc 300
|||||
Db 11333 TAAAAACAGGGATCCCGAGTAGCCTCGATGTCCCCCATTAACGGTAATATTTCAGGC 11392
|||||
QY 301 gtcgctcacactaatctttcaaaactgtcatcgcgagccgctgcccagcagattcactt 360
|||||
Db 11393 GTCGCTCACACTAATCTTTCAAACGTCTATCGGAGCGCCTGGCCAGCAGATTCACTT 11452
|||||
QY 361 aacagcgtcccgagacccctcggtccagagctctttccagcagacatttaattgaatcgg 420
|||||
Db 11453 AACAGCGTCCCGAGACCTCGTTCCGAGCTCTTTTCAGCGAGACATTTAATTGAATCGG 11512
|||||
QY 421 atgtggctggtttgcgaagctacccgctcggtgagataggtatcctctccaacgacac 478
|||||
Db 11513 ATGTGGCTCGTTGCCAGACGTACCGCCTCGGCGATAGGCATCCTCTCCAACGACAC 11570
|||||

Search completed: July 5, 2002, 10:27:10
Job time: 6432 sec

QY 361 aacagcgctccagagaccctcgctccagagctcttttccagcgagacatttaattgaatcgg 420
|||||
Db 361 aacagcgctccagagaccctcgctccagagctcttttccagcgagacatttaattgaatcgg 420
|||||
QY 421 atgtggctcgtttgcagagctcaccgctcgcgagataggcatcctctccaagacac 478
|||||
Db 421 atgtggctcgtttgcagagctcaccgctcgcgagataggcatcctctccaagacac 478
|||||
RESULT 2
AAS09962
ID AAS09962 standard; DNA; 6751 BP.
XX AC AAS09962;
XX DT 24-OCT-2001 (first entry)
XX DE Genomic DNA #2 encoding human Csx/Nkx2.5.
XX KW Csx/Nkx2.5; cardiac enhancer; cardiac cell; cardiomyocyte induction;
XX KW therapeutic; heart tissue; gene therapy; human; ds.
XX OS Homo sapiens.
XX PN WO200151006-A2.
XX PD 19-JUL-2001.
XX PF 16-JAN-2001; 2001WO-US01511.
XX PR 14-JAN-2000; 2000US-0176419.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PI Lee IW, Izumo S;
XX WPI; 2001-451809/48.
XX PT New cardiac specific cell enhancer elements, useful for specifically
PT expressing gene in cardiac cell, as earlier marker of cardiomyocyte
PT induction, e.g. for optimizing cardiomyocyte induction -
XX
XX Claim 29; Fig 4B; 66pp; English.
XX
XX The sequence represents the genomic sequence #2 of human Csx/Nkx2.5. The
CC nucleic acid is useful for specifically expressing a gene in a cardiac
CC cell, as an earlier marker of cardiomyocyte induction, e.g. for
CC optimising cardiomyocyte induction. Genes expressed in the cardiac cell-
CC specific manner are useful for the targeted expression of genes encoding
CC therapeutic proteins for the treatment of damaged heart tissue. Cardiac
CC specific enhancer elements may be used for gene therapy.
XX
XX Sequence 6751 BP; 1459 A; 2008 C; 1753 G; 1490 T; 41 other;

Query Match 100.0%; Score 478; DB 22; Length 6751;
Best Local Similarity 100.0%; Pred. No. 1.1e-175;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agagaaatattaccgattccaaaagacatagagagtgtaacagtcactgttgtt 60
|||||
Db 201 agagaaatattaccgattccaaaagacatagagagtgtaacagtcactgttgtt 260
|||||
QY 61 caaataggagaggttttttccctcttttgaacacctgacccacagagactgacag 120
|||||
Db 261 caaataggagaggttttttccctcttttgaacacctgacccacagagactgacag 320
|||||
QY 121 ttctaggaagcccttaccgaaaataggaataaaatccttgccacctgatttgaag 180
|||||
Db 321 ttctaggaagcccttaccgaaaataggaataaaatccttgccacctgatttgaag 380
|||||
QY 181 ggcaatgctaatttttttccagagctctcaaaaaaaaaaaaaaacctttac 240
|||||

Db 381 ggcaatgctaatttttttccagagctctcaaaaaaaaaaaaaaacctttac 440
QY 241 taaaaacaggagatcccgagtagcctcgatgtcccccattaaacggttaattttcaggc 300
|||||
Db 441 taaaaacaggagatcccgagtagcctcgatgtcccccattaaacggttaattttcaggc 500
|||||
QY 301 gtcgctcacactaatattttcaaaactgtcatcgagcgccgctggccagcagattcaatt 360
|||||
Db 501 gtcgctcacactaatattttcaaaactgtcatcgagcgccgctggccagcagattcaatt 560
|||||
QY 361 aacagcgctccagagaccctcgctccagagctcttttccagcgagacatttaattgaatcgg 420
|||||
Db 561 aacagcgctccagagaccctcgctccagagctcttttccagcgagacatttaattgaatcgg 620
|||||
QY 421 atgtggctcgtttgcagagctcaccgctcgcgagataggcatcctctccaagacac 478
|||||
Db 621 atgtggctcgtttgcagagctcaccgctcgcgagataggcatcctctccaagacac 678
|||||
Search completed: July 5, 2002, 10:31:08
Job time: 3910 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

AM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 08:41:18 : Search time 44.79 seconds
(without alignments)
2621.407 Million cell updates/sec

Title: US-09-761-466-6
Perfect score: 478
Sequence: : aquaaatcaatcaacaaat.....agatcctctctcaaaagaa 478

Scoring table: 1119, 1190
Gap 60, 0 : Gapext 60, 0

Searched: 84544 seqs, 12281652 residues

Word size : 50

Total number of hits satisfying chosen parameters: 0

Minimum db seq length: 0
Maximum db seq length: 200000000

Post-processing: Listing first 45 summaries

Database : ISSUED_Patents_NA:*

1:	/cqn2_5/pctodata/2/ina/5A-COMB.seq:*
2:	/cqn2_5/pctodata/2/ina/5B-COMB.seq:*
3:	/cqn2_5/pctodata/2/ina/5A-COMB.seq:*
4:	/cqn2_5/pctodata/2/ina/5B-COMB.seq:*
5:	/cqn2_5/pctodata/2/ina/PLUS-COMB.seq:*
6:	/cqn2_5/pctodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: July 5, 2002, 10:27:15
Job time: 6457 sec

GeneCore version 4.5.
Copyright (c) 1993 - 2002, Compugen, Ltd.

EM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 08:05:28 : Search time 1616.98 seconds
(without alignments)

3869.873 Million cell updates/sec

File: us-09-761-466-6

Perfect score: 478

Sequence: 1 aagaaataatcattaccgatt.....agcattccctctccaaacacac 478

Scoring table: cLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 5:

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_estr:*
- 9: qb_est1:*
- 10: qb_est2:*
- 11: qb_hic:*
- 12: qb_iss:*
- 13: em_iss_hum:*
- 14: em_iss_inv:*
- 15: em_iss_pln:*
- 16: em_iss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB ID	Description

No matches found

Search completed: July 5, 2002, 09:56:05
Job time: 6577 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuDen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 06:24:08 ; Search time 1776.56 Seconds
(without alignments)
5630.479 Million cell updates/sec

Title: US-09-761-466-6

Perfect score: 478

Sequence: 1 agagaaatcattaccgatt.....ggcatctctccaaacac 478

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 177656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*

2: gb_hfq.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_o_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

RESULT	1						
AC008412							
LOCUS	AC008412	137545 bp	DNA	linear	HTG 18-JUL-2000		
DEFINITION	Homo sapiens chromosome 5 clone C10-24H14, WIKING DRAFT SEQUENCE, 21 ordered pieces.						
ACCESSION	AC008412						
VERSION	AC008412.5	GI:5255970					
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.						
SOURCE	human.						
ORGANISM	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
REFERENCE	1 (bases 1 to 137545)						
AUTHORS	Doe Joint Genome Institute.						
TITLE	Sequencing of Human Chromosome 5						
JOURNAL	Unpublished						
REFERENCE	2 (bases 1 to 137545)						
AUTHORS	Doe Joint Genome Institute.						
TITLE	Direct Submission						
JOURNAL	Submitted (03-AUG-1999) Project on Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94596, USA						
COMMENT	On Jul 18, 2000 this sequence version replaced gi:7708438.						
	-----Genome Center						

* 116347 116446: gap of 100 bp
* 116447 131258: contig of 14812 bp in length
* 131259 131358: gap of 100 bp
* 131359 144702: contig of 13344 bp in length.

FEATURES
source

Location/Qualifiers
1. .144702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXbac-116A1"
/clone_lib="CHORI-501"
1. .75928
/note="assembly_fragment:03023
fragment_chain:1
clone_end:T7
vector_side:left"
76029. .116346
/note="assembly_fragment:03493
fragment_chain:1"
116447. .131258
/note="assembly_fragment:03524
fragment_chain:1"
131359. .144702
/note="assembly_fragment:01454
fragment_chain:1
clone_end:SP6
vector_side:right"
41517 a 33170 c 32886 g 36829 t 300 others

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT 41517 a 33170 c 32886 g 36829 t 300 others
ORIGIN

Query Match 97.5%; Score 466; DB 2; Length 144702;
Best Local Similarity 99.88; Pred. No. 8.3e-118;
Matches 477; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 agagaaatcattaccgattcacaaagacatagagagtgaacagtcactgtattgtt 60
Db 11094 AGAAAAATCATTTACCGATTACAAAAGACATAGAGAGTGAACAGTCACATCTTGT 11153
Qy 61 caaataggagagtttttttccctctttttgttaacacctgaccacagactgaac 120
Db 11154 CAAATAGGAGAGTTTTTTTCCCTCCCTTTTGTAAACACCTGACCCACAGACTGACAG 11213
Qy 121 ttctaggaagccctctaccgaaataggaataaaatcccttgccacctgattgcaag 180
Db 11214 TTCTAGGAAGCCCTTACCAGAAATAGGAATAAATCTTGCACCTTGATTGCAAG 11273
Qy 181 ggaatgctaatttttttcttccagagctctcaaaaaaataaaaccccttac 240
Db 11274 GGCAATGCTAATTTTCTTCTCCAGAGCTCTC-AAAAAATAAAACCTTAC 11332
Qy 241 taaaaacaggatcccgatgtagctgtagtgcctcattccccattaaacggttaatttcaggc 300
Db 11333 TAAAAACAGGATCCGGATGTAGCTGATGTCCTCCCATTTAACGGTAAATATTCAGCG 11392
Qy 301 gtccgctcacactaatctttcaaaactgtcatcgagcgccgtgcccagcagattcaatt 360
Db 11393 GTCCGCTCACACTAATCTTTCAAACTGTCATCGGAGCGCCTGGCCAGCAGATTCACTT 11452
Qy 361 aacagcgtccagagaccctggtccagagctcttttcagcagacatttaattgaatcgg 420
Db 11453 AACAGCGCTCCAGACCCCTCGTCCGAGCTCTTTTTCAGCGAGACATTTAATTGAATCGG 11512
Qy 421 atgtggctcgtttgcagacgtcacccgctcgggataggcatcctctccacgacac 478
Db 11513 ATGTGGCTCGTTTGGCAGACGTACCAGCTCGGCGATAGGCATCTCTCCAACGACAC 11570

RESULT 4

AF091351
LOCUS AF091351 8117 bp DNA linear ROD 23-DEC-1998
DEFINITION Mus musculus homeobox protein NKX-2.5 (Nkx-2.5) gene, complete cds.
ACCESSION AF091351

AF091351.1 GI:4050015
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 8117)
Searcy,R.D., Vincent,E.B., Liberatore,C.M. and Yutzey,K.E.
A GATA-dependent nkx-2.5 regulatory element activates early cardiac
gene expression in transgenic mice
Development 125 (22), 4461-4470 (1998)
98453405
PUBMED 9778505
REFERENCE 2 (bases 1 to 8117)
Searcy,R.D. and Yutzey,K.E.
Direct Submission
TITLE Submitted (08-SEP-1998) Molecular Cardiovascular Biology.
JOURNAL Children's Hospital Research Foundation, 3333 Burnet Avenue,
Cincinnati, OH 45229, USA
FEATURES
source
1. .8117
/organism="Mus musculus"
/strain="129/J"
/db_xref="taxon:10090"
1. .505
/gene="Nkx-2.5"
/note="GATA-dependent"
1. .5992
/gene="Nkx-2.5"
join(3057. .3629,5005. .5992)
/gene="Nkx-2.5"
/product="homeobox protein NKX-2.5"
join(3299. .3629,5005. .5630)
/gene="Nkx-2.5"
/codon_start=1
/product="homeobox protein NKX-2.5"
/protein_id="AAC97934.1"
/db_xref="GI:4050016"
/translation="MFPSPALTPTFPFVKDILNLEQOORSLASGDLSEATLAPAS
CMLAAFKPEAYSGPEAAASGLAEKLAELRAEMGAPSPKCPAPFAAPFPFGAYGDDPA
KDPADKKELCALOKAVELDKAETGAERPRARRRRKRVLFQAOVYELERRFKOOR
YLSAPERDOLASVLKLTSTQVKINFONRYKCKRQODOTLELLGPPPPARRIAVPV
LVROGKCLGDPFAAYFAIGVGLNAGYKNAIYPIPSYGGAAACSPGYSCAAYPAAPAAQ
PPAASANSFNFGVDLNTVQSPMGPOGNSVSLHGIRAW"
BASE COUNT 1831 a 2267 c 2109 g 1909 t 1 others
ORIGIN
Query Match 41.8%; Score 199.8; DB 10; Length 8117;
Best Local Similarity 74.0%; Pred. No. 1.9e-44;
Matches 328; Conservative 0; Mismatches 92; Indels 23; Gaps 5;
Qy 36 gagtgtaacagtcactgatctgtttcaaataggagaggttttttccctccctttgt 95
Db 885 GGGTGGGAAGTCATGATTTTGTTCAAATTAGAACAGTTT-CTTCTTTTCTT 939
Qy 96 aacacctgaccacacaggactgacagtcttaggaagcccttaccocgaataaggaata 155
Db 940 GACACCTGACCACACAGATTGTCAACTCT-GGAAGCCCTTATATCGGAAAA-AAAGTGATA 997
Qy 156 aatccttgccactgatttgcaagggaatgctaattttttcttctccagagcttc 215
Db 998 AATCTCCACCATGATTTGCAAGGAAACG-----TCCCTCTCAAAACTATT 1047
Qy 216 aaaaaaataaaacaccccttactaaaaacaggagatcccgatgtagcctgatgctc 275
Db 1048 TTTTAAAAAGACTTTAAACACGGATCCCGGATCGGCCCCCAATATAGTCT-----CC 1101
Qy 276 cccattaaacggttaatttcaggcgctcgctcacactaatctttcaaacgtgcatcgcg 335
Db 1102 CCAATTAACGGTAATATTTTCAGGCGTCAGCTCACACTAATCTTTTCAAACTGTTCATCGCG 1161
Qy 336 agccgcctggccagcagattcacttaaacagcgtccccagaccctcggttcccgagcttt 395


```

/note="alternatively spliced"
/product="cardiac homeobox transcription factor"
join(5100...5754,7132...8114)
/gene="Nkx2-5"
/note="alternatively spliced"
/product="cardiac homeobox transcription factor"
join(5424...5754,7132...7757)
/gene="Nkx2-5"
/function="involved in heart development"
/note="tinman homolog; Nkx family"
/codon_start=-1
/product="cardiac homeobox transcription factor"
/protein_id="AAC38875.1"
/db_xref="GI:11602838"
/translation="MFPSPALTPFVSVDKILNLEQQQSLASGDISLARLEATLAPAS
KMLAAKPEAYSGPAEAGLAEIRAEAGMPSPKPCSPAFPAAPTFYPCAYGDPPA
CDPRADKELCALQKAAVELDKAETDCAERPARRRKPRVLFSSQAQVYELERFKQQR
YLSPAERDQLASVLKLTSTQVKYWFQNRNRYCKKRQDQDTLELLGGPPPPARRIANPV
LVRDQPCGLDPAAYAGVAGVGNLGYNAIPYPYSGGAACSPGYSCAAYPAAAPPAAH
APAAANSINVFNGVDGLNTVQSPGMPQGNSTLHGTRAW"
2041 a 2336 c 2179 g 2091 t 1 others

```

Query Match	38.1%	Score 182.2	DB 10	Length 8648
Best Local Similarity	71.6%	Pred. No. 1.4e-39		
Matches 317	Conservative 0	Mismatches 103	Indels 23	Gaps 5
QY	36	gagtgtaacagtcactgatctgttccaatataggagagtggtttttttctctcccttttgt	95	
DB	3009	GGGTGGGAAGTCACTGATTTGTTCAATATAGAAGAGTTT-----CTTCCCTTTTCTT	3063	
QY	96	aacacctgacccacagactgacagtgctctaggaaagcccccttaccogaaaataaggaata	155	
DB	3064	GACACCTGACCCACAGATTTGTCAACTCT-GGAAGCCCTATATCCGAAAAA-AAAGTGATA	3121	
QY	156	aatccttgccactctgatttgcgaaggccaatgctaatttttttctctccagagctctc	215	
DB	3122	AATCCTCACCACATTTGATTGCCAAGGAACG-----TCCCTCTCAAACTATT	3171	
QY	216	aaaaaiaaaaaaiaaaacacttactaaaaacagggatccoggtatgtagctgatctcc	275	
DB	3172	TTTTTAAAAAAGACTTAAACAGGGATCCGGATGGGGGCCCAATATGTCTC-----CC	3225	
QY	276	cccaataaacgtaataatttcaggcgtccgctcacactaatcttcaaacatgtcatcgcg	335	
DB	3226	CCAATTAACGGTAATATTTCAGCGCTCAGCTCACACTAATCTTTCAAACTGTCATCGCG	3295	
QY	336	agccgcctggccagcagatcttaacttaacagcgctcccgagaccctctgtccagactctt	395	
DB	3286	ATCGCAGGGGCGACCGGATTTCATCTTTAACAGGACTCCGAGGACCTCGCCGGAAGCTGTT	3345	
QY	396	tcagcgagacattaaattgaatcgatgtgctctgttttccagacgctcacgcctcgcg	455	
DB	3346	TCAGCAAGACATTTAAATTGAATCGGATGTGCTCTGTTTGCAGACGTCACCGCTGGTG	3405	
QY	456	ataggcatctctctccaacgacac	478	
DB	3406	ATAGACATCCGTCCAACCCACAC	3428	

RESULT	7
AL591488	
LOCUS	
DEFINITION	AL591488 191494 bp DNA linear HTG 22-NOV-2001 Mus musculus chromosome 2 clone RP23-36p22, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION	AL591488
VERSION	AL591488.7 GI:17065727
KEYWORDS	HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

REFERENCE
AUTHORS      1 (sites)
TITLE        Pearce,A.
JOURNAL      Direct Submission
COMMENT      Submitted (21-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquery@sanger.ac.uk
             On Nov 25, 2001 this sequence version replaced gi:17043820.
             ----- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: humquery@sanger.ac.uk
             ----- Project Information
             Center project name: BM38P22
             ----- Summary Statistics
             Sequencing program: XGAP4; version 4.5
             Sequencing vector: plasmid; L08752; 100% of reads
             Chemistry: Dye-terminator Big Dye; 99% of reads
             Chemistry: Dye-primer Big Dye; 0% of reads
             Consensus quality: 191473 bases at least Q40
             Consensus quality: 191494 bases at least Q30
             Consensus quality: 191494 bases at least Q20
             Insert size: 191494; sum-of-contigs
             Insert size: 168060; 8.7% error; agarose-fp
             Quality coverage: 16.18x in Q20 bases; sum-of-contigs Quality
             coverage: 18.44x in Q20 bases; agarose-fp
             -----
             * NOTE: This is a 'working draft' sequence.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
             Location/Qualifiers
               1..191494
                 /organism="Mus musculus"
                 /db_xref="taxon:10090"
                 /chromosome="2"
                 /clone="RP23-36P22"
                 /clone_lib="RPC1-23"
               1..191494
                 /note="assembly fragment:03311"
BASE COUNT   49688 a 46072 c 46038 g 49696 t
ORIGIN

Query Match          9.2%; Score 44.2; DB 2; Length 191494;
Best Local Similarity 56.6%; Pred. No. 0.19;
Matches 82; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 154 taatccttgccaaccttgattgcaagggaatgctaatcttttttttttccagagctc 213
      ||| |||| ||||||| ||||||| ||||||| |||| |||| |||
Db 159235 TACAAGCTTGAGATCTTGATTCCAAAGGGCAATCTTTGCTTTCTTTTTCGCAAGGTT 159294
      ||| |||| ||||||| ||||||| ||||||| |||| |||| |||

Qy 214 tcaaaaaaaaaaaaaaaaaaacctctataaaacaggagatcccgatgtagcctcgatg 273
      ||| |||| ||||||| ||||||| ||||||| |||| |||| |||
Db 159295 ACATTAAAAAATAAAAAAATAATCAAACTACACCTACCCGATGGGTCTTAAGTCAACCC 159354
      ||| |||| ||||||| ||||||| ||||||| |||| |||| |||

Qy 274 cccccattaaacgqtaataatttcag 298
      ||| |||| ||||||| ||||||| ||||||| |||| |||| |||
Db 159355 CAGCCTTGCCACGGTACTCTCCGAG 159379
      ||| |||| ||||||| ||||||| ||||||| |||| |||| |||

RESULT 8
AC027171/c
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-404F23 map 2, LOW-PASS
SEQUENCE SAMPLING.
AC027171
AC027171.1 GI:7331490
VERSION
KEYWORDS HTG; HTGS_PHASE0.
SOURCE
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```
REFERENCE
AUTHORS
1 (bases 1 to 77483)
Birnboim, B., Linton, L., Nustbaum, C., and Lander, E.
Homo sapiens chromosome 2, clone RP11-404F23
JOURNAL
Unpublished
2 (bases 1 to 77483)
Birnboim, B., Linton, L., Nustbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barina, N., Bastien, V., Bida, F.,
Boudisaki, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collumore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gaud, D.,
Galaun, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lakocque, K., Lamazares, R., Landers, J., Lehoczeky, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marguis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
McElrath, J., Menees, L., Mihova, T., Miranda, C., Menqua, V., Morrow, J.,
Murphy, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strano-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Tridillo, J.,
Vassiliev, H., Viel, R., V.D.A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A., and Zody, N.
Direct Submission
Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
Smit, A.F.A. & Green, P. (1996-1997)
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRK
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6875
Center clone name: 404_F_23
-----
* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 784 883: contig of 783 bp in length
* 884 1715: contig of 842 bp in length
* 1716 1815: gap of 100 bp
* 1816 2635: contig of 820 bp in length
* 2636 2735: gap of 100 bp
* 2736 3523: contig of 788 bp in length
* 3524 3623: gap of 100 bp
* 3624 4447: contig of 814 bp in length
* 4438 4537: gap of 100 bp
* 4538 5363: contig of 826 bp in length
* 5364 5463: gap of 100 bp
* 5464 6265: contig of 802 bp in length
* 6266 6365: gap of 100 bp
* 6366 7189: contig of 824 bp in length
* 7190 7289: gap of 100 bp
* 7290 8104: contig of 815 bp in length
* 8105 8204: gap of 100 bp
* 8205 9008: contig of 804 bp in length
* 9009 9108: gap of 100 bp
* 9109 9939: contig of 831 bp in length
*
* 9940 10039: gap of 100 bp
* 10040 10873: contig of 834 bp in length
* 10874 10973: gap of 100 bp
* 10974 11891: contig of 828 bp in length
* 11892 12497: gap of 100 bp
* 12498 12797: gap of 100 bp
* 12798 13691: gap of 734 bp in length
* 13692 14605: contig of 814 bp in length
* 14606 15422: contig of 827 bp in length
* 15423 15533: gap of 100 bp
* 15534 16460: contig of 828 bp in length
* 16461 17278: contig of 818 bp in length
* 17279 17378: gap of 100 bp
* 17379 18214: contig of 836 bp in length
* 18215 18314: gap of 100 bp
* 18315 19152: contig of 838 bp in length
* 19153 19252: gap of 100 bp
* 19253 20088: contig of 836 bp in length
* 20089 20188: gap of 100 bp
* 20189 20963: contig of 775 bp in length
* 20964 21063: gap of 100 bp
* 21064 21894: contig of 831 bp in length
* 21895 21994: gap of 100 bp
* 21995 22769: contig of 765 bp in length
* 22769 22859: gap of 100 bp
* 22860 23645: contig of 786 bp in length
* 23646 23745: gap of 100 bp
* 23746 24541: contig of 796 bp in length
* 24542 24641: gap of 100 bp
* 24642 25447: contig of 806 bp in length
* 25448 25547: gap of 100 bp
* 25548 26373: contig of 826 bp in length
* 26374 26473: gap of 100 bp
* 26474 27287: contig of 814 bp in length
* 27288 27387: gap of 100 bp
* 27388 28210: contig of 823 bp in length
* 28211 28310: gap of 100 bp
* 28311 29107: contig of 787 bp in length
* 29108 29207: gap of 100 bp
* 29208 30038: contig of 831 bp in length
* 30039 30138: gap of 100 bp
* 30139 30966: contig of 828 bp in length
* 30967 31666: gap of 100 bp
* 31667 31872: contig of 866 bp in length
* 31873 31972: gap of 100 bp
* 31973 32772: contig of 800 bp in length
* 32773 32872: gap of 100 bp
* 32874 33672: contig of 820 bp in length
* 33673 33772: gap of 100 bp
* 33773 34587: contig of 815 bp in length
* 34588 34687: gap of 100 bp
* 34688 35506: contig of 819 bp in length
* 35507 35606: gap of 100 bp
* 35607 36418: contig of 812 bp in length
* 36419 36518: gap of 100 bp
* 36519 37340: contig of 822 bp in length
* 37341 37440: gap of 100 bp
* 37441 38266: contig of 826 bp in length
* 38267 38366: gap of 100 bp
* 38367 39191: contig of 825 bp in length
* 39192 39291: gap of 100 bp
* 39292 40075: contig of 784 bp in length
* 40076 40175: gap of 100 bp
* 40176 41007: contig of 830 bp in length
* 41008 41105: gap of 100 bp
* 41106 41919: contig of 814 bp in length
* 41920 42019: gap of 100 bp
* 42020 42822: contig of 803 bp in length
* 42823 43522: gap of 100 bp
```



```
/rpt_family="L1"
18669. .18760
/rpt_family="L1"
19660. 19723
/rpt_family="L2"
20149. .20450
/rpt_family="Alu"
20428. .20450
/rpt_family="(A)n"
21362. 21965
/rpt_family="MaLR"
21965. .22600
/rpt_family="AchoBo"
22652. .22941
/rpt_family="AchoBo"
22977. 23195
/rpt_family="AchoBo"
23455. .23809
/rpt_family="L1"
23804. .25580
/rpt_family="L1"
24583. 24621
/rpt_family="(TTTTG)n"
25594. .27618
/rpt_family="L1"
27625. .27704
/rpt_family="L1"
28570. 28691
/rpt_family="L2"
29275. .29364
/rpt_family="L1"
30438. .30466
/rpt_family="AT-rich"
31693. 31944
/rpt_family="L1"
32147. .32179
/rpt_family="(TG)n"
32463. .32634
/rpt_family="MaLR"
33036. .33221
/rpt_family="MaLR"
33325. .33364
/rpt_family="(CA)n"
33443. .33597
/rpt_family="MaLR"

Query Match      8.6%; Score 41; DB 9; Length 168698;
Best Local Similarity 61.9%; Pred. No. 1.4;
Matches 65; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 181 ggcaatgctaattttttcttccagagctctcaaaaaaaaccccttac 240
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 149868 GGTAATCTCTGTATCTACTACCTGACGAGAAAGTAAACAAAAAANAC 149927

QY 241 taaaaacaggatccccgatgtagctgcagtcctcccatataac 285
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 149928 AAAAACTGGGATCTGTCTGTAACCTGATCTAAGGGAATAACC 149972

RESULT 11
AC069242
LOCUS Homo sapiens 3 BAC RP11-554J1 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC069242
VERSION AC069242.13 GI:16596484
KEYWORDS HFG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 184664)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M.,
Holloway,C., Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S.,
Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S.,
Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E.,
Sonaake,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and
Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 184664)
Worley,K.C.
Direct Submission
Submitted (22-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 184664)
Worley,K.C.
Direct Submission
Submitted (02-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 184664)
Worley,K.C.
Direct Submission
Submitted (03-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 2, 2001 this sequence version replaced gi:16076940.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
```


AC026347	AC026347.17	GI:15920049	
VERSION	HTG.		
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 195430)		
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louleghed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S., Ogih,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Prabakaran,I., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,A., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and Gibbs,R.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 195430)		
AUTHORS	Worley,K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
REFERENCE	3 (bases 1 to 195430)		
AUTHORS	Worley,K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
COMMENT	On Oct 4, 2001 this sequence version replaced gi:15809081. INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu		
CLONE LENGTH:	This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.		

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.
Location/Qualifiers
1. 195430
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-362A9"
500..521
/rpt_family="AT_rich"
1223..1252
/rpt_family="AT_rich"
3700..3720
/rpt_family="AT_rich"
complement(4526..4693)
/rpt_family="MLT13"
4805..4857
/rpt_family="T-rich"
complement(5206..5501)
/rpt_family="AluX"
complement(5690..5930)
/rpt_family="MIR"
7018..7041
/rpt_family="TTTTG)n"
complement(8516..8653)
/rpt_family="MIR"
8909..8937
/rpt_family="AT_rich"
complement(9215..9519)
/rpt_family="AluSg"
9789..9823
/rpt_family="AT_rich"
10126..10425
/rpt_family="AluYa5"
complement(11240..11535)
/rpt_family="AluSg"
11572..11790
/rpt_family="MIR"
complement(12590..12887)
/rpt_family="AluSp"
14059..14080
/rpt_family="(CA)n"
14081..14115
/rpt_family="(TC)n"
15047..15081
/rpt_family="(TTC)n"
15242..15269

KEYWORDS

SOURCE
ORGANISM

HTG.

Caenorhabditis elegans.

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1. (bases 1 to 48488)

The C. elegans Sequencing Consortium.

Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium

Science 282 (5396), 2012-2018 (1998)

99069613

2. (bases 1 to 48488)

Woesner, J., Graves, T. and Keppler, D.

The sequence of C. elegans cosmid Y44E3A

Unpublished

3. (bases 1 to 48488)

Waterston, R.

Direct Submission

Unpublished

4. (bases 1 to 48488)

Waterston, R.

Submitted (15-NOV-1998) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

5. (bases 1 to 48488)

Waterston, R.

Direct Submission

Submitted (05-MAR-1999) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6. (bases 1 to 48488)

Waterston, R.

Direct Submission

Submitted (30-JUN-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

7. (bases 1 to 48488)

Waterston, R.

Direct Submission

Submitted (03-JUL-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

8. (bases 1 to 48488)

Waterston, R.

Direct Submission

Submitted (16-NOV-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

Louis, MO 63110, USA

On Jun 30, 2001 this sequence version replaced gi:3886084.

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RQ, England

email: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE:

This sequence may not be the entire insert of this clone.

It may be shorter because we only sequence overlapping sections

once, or longer because we provide a small overlap between

neighboring submissions.

This sequence was finished as follows unless otherwise noted: all

regions were double stranded, sequenced with an alternate chemistry

or covered by high quality data (i.e., phred quality >= 30); an

attempt was made to resolve all sequencing problems, such as

compressions and repeats; all regions were covered by sequence from

more than one ml3 subclone.

NOTES:

Coding sequences below are predicted from computer analysis, using

the program GeneFinder (P. Green and L. Hillier, ms in preparation).

Location/Qualifiers

FEATURES

.

source

1. 48488

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="I"

/clone="Y44E3A"

complement(51..5319)

/gene="Y44E3A.2"

/note="ace-2"

complement(join(51..175,348..508,1022..1195,1721..2145,

3031..3267,3314..3542,3798..3973,4078..4281,5161..5319))

/gene="Y44E3A.2"

/note="C. elegans acetylcholinesterase (ACED-2)

(GB:AF025378); similar to carboxylesterases (Pfam:

PF00135, Score=363.2, E=4.9e-106, N=1); coded for by the

following C. elegans cDNAs: AF025378"

/codon_start=1

/product="Hypothetical protein Y44E3A.2"

/protein_id="AAC78228.2"

/db_xref="GI:14578252"

/translation="MRAPVIGRHLYHVFQFALVTLFIVRRIRPSIVRGDHHVHTP

LGTIRGVQTFDGAQVSAFLGVPAKPIGSRFRKMAEMIDRWSELEARTLAKTCYL

TIDSAFPQFCAEMWNPAGISEDCLNNIWPEDHDCVMVWLYGGFFSCTPSIDL

YGSVFAAKDHTIVVNYRLGPFGLYFGDDSPIQNGMLDQOLALRWVHENGICAF

GDSRVTLTFGESAGSASTTAHLFAPNSHKYFRNLIAGSGSIINWSASATPTMLDLS

FRLAKVNCSSPDMAIVKLSRVPALHQAEDNISGDIQPMTEAFVVPVSSDANFF

QGDVFOKLANKQFKKDVNIIFGSVKDEGYMLPYMSLPKYFAFNHTISAEDPHNRA

LITRDHYEESMRAMPYFAGSKVLVNAFMNSYEHVSTSNVPEERYDGVARFLGDLFF

TCSLIDFADLLSUNIFGVNYMYFTYRSSANPWPKWGMVGHYELEYAFGOPYWRPHL

YDTHLEDERKRLSIIIMQIWANFANTGRTDSFWQYKIERKAELELGGTTTQGGKRRIL

SDVHGFGFRMIDEAKAFVKQKANDCTRTRKASTEDLTSSSTYUFSIIVYSILI

SYISL"

complement(15909..16429)

/gene="Y44E3A.3"

complement(join(15909..16140,16338..16429))

/gene="Y44E3A.3"

/note="contains similarity to thioredoxins (Pfam: PF00085,

Score=99.8, E=9.3e-29, N=1)"

/codon_start=1

/product="Hypothetical protein Y44E3A.3"

/protein_id="AAC78230.1"

/db_xref="GI:3886088"

/translation="MSIAIKDKDEEFTFAFKTKTPQVILFTFASWCGPCQMIKRVEE

LAAEHKDRLSILKIDVDECDGVGEINSMPTELLIVDGIKKDQFSGANNTKFEEMV

KAALQ"

complement(17707..21817)

/gene="Y44E3A.4"

complement(join(17707..17876,18825..19047,20000..20206,

20823..20984,21120..21160,21280..21397,21513..21677,

21779..21817))

/gene="Y44E3A.4"

/codon_start=1

/product="Hypothetical protein Y44E3A.4"

/protein_id="AAC78229.1"

/db_xref="GI:3886087"

/translation="MKDKIFKIKTKKADPPPTIAPAAPSAPAPASQALNPTSTPT

MPALAPSSALKNSFAALQEKMSRKILQKKRNKQFFLGLFLKFNKNSLKRKNEK

KPKNQKLSFKIDSIFGEFKNFFSKIEYNKNCSEFSENHSEKQFSLFFCFENFI

VLELNMSNVLKITLNLKVAPEKVASIOKPDORRSTGESTTGEPSIDNSADOVA

ELQHTKNRARPKNRPMVMSVMNRNSDESDEPNGLSSPTISITVNSASMEVPAPA

KEISIPNSATSPFKSSSIAPPSASKPALRPIPSNLADDSSTSSAISRLKASPD

SQKSPNLPSPDYSVRAEYDALLDRFRALESRLAIEKIRF"

complement(24297..28465)

/gene="Y44E3A.5"

complement(join(24297..24584,25732..25935,26547..26890,

27694..27796,28448..28465))

/gene="Y44E3A.5"

/note="contains similarity to Src homology domain 3 (Pfam:

PF00018, Score=116.8, E=4e-31, N=2); coded for by the

following C. elegans cDNAs: yk341h3.5"

/codon_start=1

/product="Hypothetical protein Y44E3A.5"

/protein_id="AAC78231.1"

Claim 21; Fig 5c; 66pp; English.

The sequence represents the coding sequence of cardiac enhancer hCscx/Nkx2.5 homology domain B. The nucleic acid is useful for specifically expressing a gene in a cardiac cell, as an earlier marker of cardiomyocyte induction, e.g. for optimising cardiomyocyte induction. Genes expressed in the cardiac cell-specific manner are useful for the targeted expression of genes encoding therapeutic proteins for the treatment of damaged heart tissue. Cardiac specific enhancer elements may be used for gene therapy.

Sequence 478 BP: 138 A; 129 C; 89 G; 122 T; 0 other;

14-JAN-2000; 2000US-0176419.

(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

Lee IW, Izumo S;

WPI; 2001-451809/48.

New cardiac specific cell enhancer elements, useful for specifically expressing gene in cardiac cell, as earlier marker of cardiomyocyte induction, e.g. for optimizing cardiomyocyte induction -

Claim 29; Fig 4B; 66pp; English.

The sequence represents the genomic sequence #2 of human Csx/Wkx2.5. The nucleic acid is useful for specifically expressing a gene in a cardiac cell, as an earlier marker of cardiomyocyte induction, e.g. for optimizing cardiomyocyte induction. Genes expressed in the cardiac cell-specific manner are useful for the targeted expression of genes encoding therapeutic proteins for the treatment of damaged heart tissue. Cardiac specific enhancer elements may be used for gene therapy.

Sequence 6751 BP; 1459 A; 2008 C; 1753 G; 1490 T; 41 other;

Query Match 100.0%; Score 478; DB 22; Length 6751;
 Best Local Similarity 100.0%; Pred. No. 5.1e-122;
 Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agagaaatcattaccgattccaaaagagatagagagttaacagtcactgattctgtt 60
 |||||
 Db 201 agagaaatcattaccgattccaaaagagatagagagttaacagtcactgattctgtt 260
 |||||
 Qy 61 caaataggagaggttttttccctccctttttaaaccctgacccacagagctgacag 120
 |||||
 Db 261 caaataggagaggttttttccctccctttttaaaccctgacccacagagctgacag 320
 |||||
 Qy 121 ttctaggaagcccttaccgaaataggaataataatcctgccactgatttgcaag 180
 |||||
 Db 321 ttctaggaagcccttaccgaaataggaataataatcctgccactgatttgcaag 380
 |||||
 Qy 181 gcaatgctaattttttctccagagctctcaaaaaaataaaaaaaccttac 240
 |||||
 Db 381 gcaatgctaattttttctccagagctctcaaaaaaataaaaaaaccttac 440
 |||||
 Qy 241 taaaaacaggatcccgagtagcctcgatgctcccccattaaacggttaatttcaggc 300
 |||||
 Db 441 taaaaacaggatcccgagtagcctcgatgctcccccattaaacggttaatttcaggc 500
 |||||
 Qy 301 gtccgctcaactaatotttcaaaactgcatcgagccgcctggccagagattcactt 360
 |||||
 Db 501 gtccgctcaactaatotttcaaaactgcatcgagccgcctggccagagattcactt 560
 |||||
 Qy 361 aacagcgtccagagaccctcgttccagagctctttcagcgagacatttaattgaatcg 420
 |||||
 Db 561 aacagcgtccagagaccctcgttccagagctcttttcagcgagacatttaattgaatcg 620
 |||||
 Qy 421 atgtggctcgtttgccagaagctcaccgcctcgcgataggcattcctccaacgacac 478
 |||||
 Db 621 atgtggctcgtttgccagaagctcaccgcctcgcgataggcattcctccaacgacac 678
 |||||

RESULT 3
 AAC59534
 ID AAC59534 standard; cDNA; 1086 BP.
 XX
 AC AAC59534;
 XX
 XX 15-FEB-2001 (first entry)
 XX
 DE Human secreted protein cDNA sequence #28.
 XX
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;


```

PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457723/49.
DR
XX Isolated polypeptide for treating, preventing and/or prognosing
PT respiratory disorders related to the lung including lung cancers and
PT also for testing and detection e.g. diagnosis -
XX
XX Claim 1; SEQ ID NO 362; 507pp; English.
PS
XX Sequences AAS29931-AA330164 represent genomic DNA molecules, which encode
CC the lung antigen polypeptides of the invention. Lung antigen polypeptides
CC and their associated polynucleotides are useful in the diagnosis,
CC treatment and prevention of various types of disorders in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by detecting the presence or
CC absence of a mutation in a lung antigen polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
Query Match      8.2%; Score 39; DB 22; Length 3276;
Best Local Similarity 68.4%; Pred. No. 0.85;
Matches 54; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 177 caaggcgaatgctaatttttcttcctccagcgctctcaaaaaaaaaaaaaaaccc 236
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2059 catggacaaatgaatttatcttcttaccgttgtaaaaactataaaaaaaccc 2118
QY 237 ttactaaaaacagggatccc 255
   | | ||||| | |||
Db 2119 taataaaaaaaaagaagctcc 2137
   | | ||||| | |||
RESULT 6
ID AAI93585/c
ID AAI93585 standard; cDNA; 397 BP.
XX
XX AAI93585;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 13645.
DE
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;

```

KW	nervous system disorders; arthritis; inflammation; ss.		
OS	Homo sapiens.		
XX			
XX	WO200164835-A2.		
XX			
PD	07-SEP-2001.		
XX			
XX	26-FEB-2001; 2001WO-US04927.		
XX			
PF	28-FEB-2000; 2000US-0515126.		
XX			
PR	18-MAY-2000; 2000US-0577409.		
XX			
XX	(HYSE-) HYSEQ INC.		
PA			
XX			
PI	Tang YT, Liu C, Drmanac RT;		
XX			
DR	WPI; 2001-514838/56.		
XX			
DR	P-PSDB; AAO13654.		
XX			
PT	Isolated nucleic acids and polypeptides, useful for preventing		
PT	diagnosing and treating e.g. leukaemia, inflammation and immune		
XX	disorders -		
XX			
PS	Claim 1; SEQ ID NO 13645; 1399pp + Sequence Listing; English.		
XX			
CC	The invention relates to human polynucleotides (AAI79941-AAI93841) and		
CC	the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to		
CC	cytokine, cell proliferation or cell differentiation or which may induce		
CC	production of other cytokines in other cell populations. The		
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or		
CC	peptide therapy. The polypeptides have various cytokine-like activities,		
CC	e.g. stem cell growth factor activity, haematopoiesis regulating		
CC	activity, tissue growth factor activity, immunomodulatory activity and		
CC	activin/inhibin activity and may be useful in the diagnosis and/or		
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and		
CC	inflammation.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pat_sequences.		
XX			
SQ	Sequence 397 BP; 115 A; 95 C; 72 G; 115 T; 0 other;		
Query Match 8.0%; Score 38.4; DB 22; Length 397;			
Best Local Similarity 56.2%; Pred. No. 0.57; 56; Indels 0; Gaps 0;			
Matches 72; Conservative 0; Mismatches 0; Gaps 0;			
Qy	121 ttctaggagagcccttaccggaaataaggaaataatccttgcaccttgattgcaag 180		
Db	208 TTTTGTGGGCCCCCAAAAAAAAAAAAAAAAAAAACCCCCCTTTATTTTGGGGG 149		
Qy	181 ggcaatgctaattttttcttccagagctctcaaaaaaaaaaaaaaaccttac 240		
Db	148 GGGGCCCTTTTTTTTTTTTTTTTTTTCCGCCCAACAAAAAACAACAAAAA 89		
Qy	241 taaaaaca 248		
Db	88 AAAAAAAA 81		
RESULT	7		
AAI91430			
ID	AAI91430 standard; cDNA; 403 BP.		
XX			
XX	AAI91430;		
XX			
DF	06-NOV-2001 (first entry)		
XX			
DE	Human polynucleotide SEQ ID NO 11490.		
XX			
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;		
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;		

PR	02-OCT-2000;	2000US-0237040;
PR	13-OCT-2000;	2000US-0239937;
PR	13-OCT-2000;	2000US-0239937;
PR	20-OCT-2000;	2000US-0240960;
PR	20-OCT-2000;	2000US-0241221;
PR	20-OCT-2000;	2000US-0241785;
PR	20-OCT-2000;	2000US-0241786;
PR	20-OCT-2000;	2000US-0241787;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241809;
PR	01-NOV-2000;	2000US-0241826;
PR	01-NOV-2000;	2000US-0244617;
PR	08-NOV-2000;	2000US-0246474;
PR	08-NOV-2000;	2000US-0246475;
PR	08-NOV-2000;	2000US-0246476;
PR	08-NOV-2000;	2000US-0246477;
PR	08-NOV-2000;	2000US-0246478;
PR	08-NOV-2000;	2000US-0246523;
PR	08-NOV-2000;	2000US-0246524;
PR	08-NOV-2000;	2000US-0246525;
PR	08-NOV-2000;	2000US-0246526;
PR	08-NOV-2000;	2000US-0246527;
PR	08-NOV-2000;	2000US-0246528;
PR	08-NOV-2000;	2000US-0246532;
PR	08-NOV-2000;	2000US-0246609;
PR	08-NOV-2000;	2000US-0246610;
PR	08-NOV-2000;	2000US-0246611;
PR	08-NOV-2000;	2000US-0246613;
PR	17-NOV-2000;	2000US-0249207;
PR	17-NOV-2000;	2000US-0249208;
PR	17-NOV-2000;	2000US-0249209;
PR	17-NOV-2000;	2000US-0249210;
PR	17-NOV-2000;	2000US-0249211;
PR	17-NOV-2000;	2000US-0249212;
PR	17-NOV-2000;	2000US-0249213;
PR	17-NOV-2000;	2000US-0249214;
PR	17-NOV-2000;	2000US-0249215;
PR	17-NOV-2000;	2000US-0249216;
PR	17-NOV-2000;	2000US-0249217;
PR	17-NOV-2000;	2000US-0249218;
PR	17-NOV-2000;	2000US-0249244;
PR	17-NOV-2000;	2000US-0249245;
PR	17-NOV-2000;	2000US-0249264;
PR	17-NOV-2000;	2000US-0249265;
PR	17-NOV-2000;	2000US-0249297;
PR	17-NOV-2000;	2000US-0249299;
PR	17-NOV-2000;	2000US-0249300;
PR	01-DEC-2000;	2000US-0250160;
PR	01-DEC-2000;	2000US-0250391;
PR	05-DEC-2000;	2000US-0251030;
PR	05-DEC-2000;	2000US-0251988;
PR	06-DEC-2000;	2000US-0256719;
PR	06-DEC-2000;	2000US-0251479;
PR	08-DEC-2000;	2000US-0251856;
PR	08-DEC-2000;	2000US-0251868;
PR	08-DEC-2000;	2000US-0251869;
PR	08-DEC-2000;	2000US-0251989;
PR	08-DEC-2000;	2000US-0251990;
PR	11-DEC-2000;	2000US-0254097;
PR	05-JAN-2001;	2001US-0259678;
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC
XX		
PI	Rosen CA,	Barash SC, Ruben
XX		
DR	WPI;	2001-483426/52.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK34942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention.

Sequence 10740 BP: 3239 A: 1965 C: 2229 G: 3307 T: 0 other

XX 15-MAR-2001; 2001WO-EP02945.
PF 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-602751/68.
XX Designing primers and probes for analysing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle -
XX Claim 1; SEQ ID NO 185; 28pp; English.
XX Sequences AAS45296-AAS4520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers.
XX Sequence 11047 BP; 3237 A; 248 C; 2358 G; 5204 T; 0 other;

Query Match 7.9%; Score 37.8; DB 22; Length 11047;
Best Local Similarity 51.1%; Pred. No. 2.9;
Matches 115; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
Qy 24 aaagacatagagagtgaacagtcactgtctgttcacaaataggagaggttttttcc 83
Db 7544 AAAAAAAAAAAAAATATACATCAATTAATTAATAAAAAATATTTATCTTTAAACC 7485
Qy 84 ttccctttttgttaa---cacctgacccacagagactgacagttctaggagcccccctacc 140
Db 7484 TCCATTATTACGAATTTCAATTCACCAATAAATCAATAATATTATAGGCACCTACTTCT 7425
Qy 141 cgaataataggaaataaactcttgccaccttgattgcaagggaatgctaatttttct 200
Db 7424 TCTAAATCTTCATTAATAACTACTAACCATATACTAAATCGCTATTTCCAAATTTAAAT 7365
Qy 201 ttctccagagctctcaaaaaaaaaaaaaaaaaaaccttactaaaa 245
Db 7364 ATTTCTCACTATATAAAAAAAAAAAAAATAAATAAATACTTAA 7320

RESULT 14
ABL33985/C
ID ABL33985 standard; DNA; 11047 BP.
XX ABL33985;
XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 1958.
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX Homo sapiens.
OS WO200200928-A2.
PN 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP07537.
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
DR Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX Claim 1; SEQ ID NO 1958; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulceration bowel
CC diseases. The present sequence is a gene of the invention.
XX Sequence 11047 BP; 3237 A; 248 C; 2358 G; 5204 T; 0 other;

Query Match 7.9%; Score 37.8; DB 24; Length 11047;
Best Local Similarity 51.1%; Pred. No. 2.9;
Matches 115; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
Qy 24 aaagacatagagagtgaacagtcactgtctgttcacaaataggagaggttttttcc 83
Db 7544 AAAAAAAAAAAAAATATACATCAATTAATTAATAAAAAATATTTATCTTTAAACC 7485
Qy 84 ttccctttttgttaa---cacctgacccacagagactgacagttctaggagcccccctacc 140
Db 7484 TCCATTATTACGAATTTCAATTCACCAATAAATCAATAATATTATAGGCACCTACTTCT 7425
Qy 141 cgaataataggaaataaactcttgccaccttgattgcaagggaatgctaatttttct 200
Db 7424 TCTAAATCTTCATTAATAACTACTAACCATATACTAAATCGCTATTTCCAAATTTAAAT 7365
Qy 201 ttctccagagctctcaaaaaaaaaaaaaaaaaaaccttactaaaa 245
Db 7364 ATTTCTCACTATATAAAAAAAAAAAAAATAAATAAATACTTAA 7320

RESULT 15
AAH76837
ID AAH76837 standard; cDNA; 5027 BP.
XX AAH76837;
XX 14-DEC-2001 (first entry)
XX Human vesicular transport-related protein 11-encoding cDNA.

KW Human; vesicular transport-related protein 11; recombinant production;
KW malignant tumor; cancer; blood disease; HIV infection;
KW human immunodeficiency virus; immune disorder; inflammatory condition;
KW cytostatic; anti-HIV; antiinflammatory; immunomodulator; ss.
XX
CS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 4448..4756
FT /*tag= a
FI /product= "Human vesicular transport-related protein 11"
XX
PN W:200172815-A1.
XX
HD 04-OCT-2001.
XX
PF 26-MAR-2001; 2001WO-CN00493.
XX
PE 27-MAR-2000; 2000CN-0115175.
XX
PA (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
PI Mao Y, Xie Y;
XX
DR WFI: 2001-602851/68.
LR P-PSDB: AAG66760.
XX
XX Human vesicular transport-related protein 11 and encoded
PI polynucleotide, applicable in diagnosis and treatment of malignant
PI tumor, hemopathy, HIV infection, immunological diseases and various
PI inflammations.
XX
PS Claim 6; Page 29-32; 36pp; Chinese.
XX
CC This sequence represents cDNA encoding human vesicular transport-related
CC protein 11. The protein has a molecular weight of 11 kD. The invention
CC relates to human vesicular transport-related protein 11 (AAG66760),
CC nucleic acids encoding it (AAG76837), and a method for the recombinant
CC production of vesicular transport-related protein 11. The present
CC invention additionally discloses an antagonist of vesicular transport-
CC related protein 11 for therapeutic use, and an antibody which
CC specifically binds to vesicular transport-related protein 11. Vesicular
CC transport-related protein 11, and nucleotides which encode it may be
CC used for treating a variety of diseases, such as malignant tumors,
CC blood diseases, HIV (human immunodeficiency virus) infection, immune
CC disorders and inflammatory conditions. The protein may also be used to
CC screen for modulators of its activity or for peptide fingerprinting
CC identification. The polynucleotide can be used as a primer for nucleic
CC acid amplification reactions or as a probe for hybridisation reactions,
CC or in producing gene chips or microarrays.
XX
SQ Sequence 5027 BP; 1811 A; 760 C; 735 G; 1721 T; 0 other;

Query Match 7.8%; Score 37.4; DB 22; Length 5027;
Best Local Similarity 46.7%; Pred. No. 2.8;
Matches 119; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
G 30 catagaagaagtgaacagtcactgtctgttccaaatagaagagag:ttttttctctcct 89
D 1618 cgttaacatttaagtgtgaaatttgtattttctataagagattgtcattctactact 1677
G 90 tttttataacacgtgaccacagagagactgacagttcttagaagaagcccttaccggaataat 149
D 1678 ttttgtatgtcgtatgagtagtataatgattcaagataatttgcacttagattt 1717
G 150 aaataaatactcttgcaccttgatttgcagaaggaactgctaaattttttttcttccaga 209
D 1738 ggaatatagtcttcttttatacatataaagaagataaatacacaatacttcttcaaguc 1747
G 210 gctctcaaaaaaataaaaaaaccttactaataaaaaaggaaccccaatgtatgcctca 269
D 1798 ataaacacacaaagaagaacaccttttaggaaactaaatgataatcttctctttt 1857


```
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/14073
; FILING DATE: 08-DEC-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-072 (IMI-041)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1059 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 63..848
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 150..848
; PCT-US94-14073-1

Query Match 6.9%; Score 33; DB 5; Length 1059;
Best Local Similarity 57.1%; Pred. No. 1.5;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 144 aaatagagaaataatccttgcacaccttgattgcaagggaatgctaatctttttcttc 203
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 951 ATATTGAGCTGAAACAAACAAATCATGATTGATAGGATTTCCTGTTGTTTCAATT 1010
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 204 tccagagctctcaaaaaaataaaaaaaaccccttactaaaaa 248
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1011 TAAAAAATCCAAAATAAATAATAATAATCATTTCAACAAAAA 1055
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-08-487-810-1
; Sequence 1, Application US/08487810
; Patent No. 5618695
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA ENCODING HEM-1, A GENE EXPRESSED BY
; TITLE OF INVENTION: SCLEROSING HEMANGIOMA CELLS
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1146 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-487-810-1

Query Match 6.9%; Score 32.8; DB 1; Length 1146;
Best Local Similarity 59.8%; Pred. No. 1.8;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 143 aaatagagaaataatccttgcacaccttgattgcaagggaatgctaatctttttcttc 202
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1054 AAAATCGCAGTTAAGCATGAACACATTGAATTGAAAAAGAACTCTGTAGTTTGAGATT 1113
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 203 ctccagagctctcaaaaaaataaaaaaa 234
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1114 TGCCATACAGCAAGAAAAAANAANAANA 1145
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-08-553-367A-5
; Sequence 5, Application US/08553367A
; Patent No. 5939539
; GENERAL INFORMATION:
; APPLICANT: Theodor LANGE et al.
; TITLE OF INVENTION: GA 20-OXIDASE GENE SEQUENCES (As Amended)
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,367A
; FILING DATE: No. 5939539ember 27, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: 49/FD4.5WZ
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1490 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; IMMEDIATE SOURCE:
; CLONE: PAT2353
; US-08-553-367A-5

Query Match 6.9%; Score 32.8; DB 2; Length 1490;
Best Local Similarity 59.8%; Pred. No. 2;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```


Search completed: July 5, 2002, 08:06:25
Job time: 6517 sec

Genfore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 06:11:43 : Search time 1604.3 Seconds
(without alignments)
4021.408 Million cell updates/sec

Title: us-09-761-466-6

Perfect score: 478
Sequence: 1 aagaaatcattaccgatt.....gacatctctccaaacac 478

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 1

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estim1.*
2: em_estim2.*
3: em_estim3.*
4: em_estim4.*
5: em_estim5.*
6: em_estim6.*
7: em_estim7.*
8: em_estim8.*
9: em_estim9.*
10: em_estim10.*
11: em_estim11.*
12: em_estim12.*
13: em_estim13.*
14: em_estim14.*
15: em_estim15.*
16: em_estim16.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Res Hit No.	Score	Query Match	Length	DB ID	Description
C 1	45.6	9.5	1101	12	CNS016PM
C 2	43.6	9.1	838	16	BG781725
C 3	43.2	9.0	1244	12	CNS01617
C 4	41.8	8.7	453	12	CNS0145V
C 5	41.8	8.7	1101	12	CNS0019R
C 6	41.6	8.7	289	9	A1947755
C 7	41.6	8.7	381	9	A1947756
C 8	41	8.6	791	12	CNS041PQ
C 9	40.8	8.5	829	9	AL514053
C 10	40.8	8.5	1201	12	CNS01050
C 11	40.6	8.5	367	10	B1510842
C 12	40.2	8.4	859	12	CNS004YY
C 13	40.2	8.4	980	12	CNS000UG
C 14	40.2	8.4	986	12	CNS0067R
C 15	40.2	8.4	1101	12	CNS00AY1
C 16	40	8.4	267	10	C91850
C 17	40	8.4	343	9	AL355572

C 18	40	8.4	1101	12	CNS01471
C 19	40	8.4	1543	16	B152111
C 20	39.8	8.3	328	12	AQ088888
C 21	39.8	8.3	697	12	CNS151S
C 22	39.8	8.3	906	12	CNS167W
C 23	39.6	8.3	450	12	AQ54261R
C 24	39.6	8.3	529	10	BF288712
C 25	39.6	8.3	540	12	AQ202762
C 26	39.4	8.2	1101	12	CNS00HPS
C 27	39.2	8.2	1019	10	BF741517
C 28	39	8.2	520	9	A170147R
C 29	39	8.2	1101	12	CNS01493
C 30	39	8.2	1201	12	CNS017JH
C 31	38.8	8.1	352	9	AL500494
C 32	38.8	8.1	466	9	AL365934
C 33	38.8	8.1	902	12	CNS0088M
C 34	38.6	8.1	186	9	AL37121*
C 35	38.6	8.1	253	9	A1829472
C 36	38.4	8.0	653	10	BG247756
C 37	38.4	8.0	880	9	AL521664
C 38	38.4	8.0	1101	12	CNS0160*
C 39	38.4	8.0	1101	12	CNS0176S
C 40	38.2	8.0	268	16	BF76344*
C 41	38.2	8.0	272	10	BM17414
C 42	38.2	8.0	295	10	BF15013*
C 43	38.2	8.0	422	9	AW64743*
C 44	38.2	8.0	593	9	AW97044*
C 45	38.2	8.0	600	9	AL47253*

ALIGNMENTS

RESULTS :
CNS016PM
1000S
DEFINITION
Prosophila melanogaster genomic survey sequence (1 to 1101)
BAC16614 of Drosophila melanogaster (1 to 1101)
fly), genomic survey sequence.

ALIGNMENT
AL07044
AL07044.1 31:56.2[57

VERSION
fruit fly.

KEYWORDS
GSS.

SOURCE
ORGANISM

Prosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Tephritidae; Tephritidae; Tephritidae;

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (28-JUL-1999) Genoscope - Centre National de Sequenciation

RP191 91006 EVERY CODEX - PFANKE (E-mail: seph@genoscope.cns.fr)

Web: www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a

collaboration with the European Institute of Science and Technology (EIT) -

Library (Dros BAC) was made by Alain Bouchard at EIT (Centre

Ecole du Polytechnique) with funding provided by a MRC

project grant. The DNA was prepared from cell lines by Alain Bouchard

and Genevieve Payan. It has been constructed in the vector

pBeloBAC11.

Location/Qualities

1101

Organism="Prosophila melanogaster"

/Eukaryota="Insecta"

/Drosophila="Tephritidae"

/clone="BAC16614"

/note="end: 10"

BASE COUNT

333 a 158 c 89 t 277 g 163

ORIGIN

AL365572

AL365572

AL365572

AL365572

AL365572


```

Qy 118 cagttcttaggaagcccccctaccgaaataaggaaataatccttggccacctgtattgc 177
      ||| || || || || || || || || || || || || || || || || || || ||
Db 166 CAGTTTACGGGTGTCGCCCATCCACCATGATAATATAGTTTGCCTAATTTTCTC 225
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 178 aagggaatgctaattttttttcttccagagactctcaaaaaaaaaaaaaaaaaac 237
      || || || || || || || || || || || || || || || || || || || ||
Db 226 AGACGCTTCCCAATCGAAGCTTTTCTGCCATGCTTAAAAAATTTTAAAAA 285
      || || || || || || || || || || || || || || || || || || || ||

RESULT 7
AL947756
LOCUS
DEFINITION
  603027H09.x1 603 - stressed root cDNA library from Wang/Bohnert lab
  Zea mays cDNA, mRNA sequence.
ACCESSION
  AL947756
VERSION
  AL947756.1 GI:5740066
KEYWORDS
  EST.
SOURCE
  Zea mays.
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
  clade; Panicoideae; Andropogoneae; Zea.
  1 (bases 1 to 381)
REFERENCE
  Walbot,V.
  Maize ESTs from various cDNA libraries sequenced at Stanford
  University
  Unpublished (1999)
  Contact: Walbot V
  Department of Biological Sciences
  Stanford University
  855 California Ave, Palo Alto, CA 94304, USA
  Tel: 650 723 2227
  Fax: 650 725 8221
  Email: walbot@stanford.edu
  Plate: 603027 row: H column: 09.
FEATURES
  source
    1..381
    /organism="Zea mays"
    /cultivar="B73"
    /db_xref="taxon:4577"
    /clone_lib="603 - stressed root cDNA library from
    Wang/Bohnert lab"
    /tissue_type="seedling"
    /dev_stage="salt stress"
    /lab_host="E. coli XL Gold"
    /note="Organ: root; Vector: pBluescriptII SK(+); XR;
    Seedling stressed root cDNA library from Wang/Bohnert lab"
  BASE COUNT 122 a 72 c 84 g 103 t
  ORIGIN

  Query Match 8.7%; Score 41.6; DB 9; Length 381;
  Best Local Similarity 59.2%; Pred. No. 3.7e+02;
  Matches 71; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 118 cagttcttaggaagcccccctaccgaaataaggaaataatccttggccacctgtattgc 177
      ||| || || || || || || || || || || || || || || || || || || ||
Db 258 CAGTTTACGGGTGTCGCCCATCCACCATGATAATATAGTTTGCCTAATTTTCTC 317
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 178 aagggaatgctaattttttttcttccagagactctcaaaaaaaaaaaaaaaaaac 237
      || || || || || || || || || || || || || || || || || || || ||
Db 318 AGACGCTTCCCAATCGAAGCTTTTCTGCCATGCTTAAAAAATTTTAAAAA 377
      || || || || || || || || || || || || || || || || || || || ||

RESULT 8
CNS04DPQ/c
LOCUS
DEFINITION
  791 bp DNA linear GSS 21-MAY-2000
  Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
  103G11 of library G from Tetraodon nigroviridis, genomic survey
  sequence.
ACCESSION
  AL286055
VERSION
  AL286055.1 GI:8024497
KEYWORDS
  GSS; genome survey sequence.

```

```

SOURCE
ORGANISM
  Tetraodon nigroviridis.
  Tetraodon nigroviridis
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
  Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
  Tetraodontidae; Tetraodon.
  1 (bases 1 to 791)
REFERENCE
  Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
  Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
  Weissenbach,J.
  Characterization and repeat analysis of the compact genome of the
  freshwater pufferfish Tetraodon nigroviridis
  Unpublished
  2 (bases 1 to 791)
REFERENCE
  Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
  Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
  Saurin,W. and Weissenbach,J.
  Human gene number estimate provided by genome wide analysis using
  Tetraodon nigroviridis DNA sequence
  Unpublished
  3 (bases 1 to 791)
REFERENCE
  Genoscope.
  Direct Submission
  Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
  This sequence is a single read and was generated as part of a large
  scale clone-end sequencing project of the Tetraodon nigroviridis
  genome. For more information, please take a look at
  http://www.genoscope.cns.fr/Tetraodon.
FEATURES
  source
    1..791
    /organism="Tetraodon nigroviridis"
    /db_xref="taxon:99883"
    /clone="103G11"
    /clone_lib="G"
    /note="Genoscope sequence ID : COBG103AD06SP1-end ;
    PUC-ori"
  BASE COUNT 208 a 146 c 144 g 273 t 20 others
  ORIGIN

  Query Match 8.6%; Score 41; DB 12; Length 791;
  Best Local Similarity 48.6%; Pred. No. 3.2e+02;
  Matches 88; Conservative 8; Mismatches 85; Indels 0; Gaps 0;

Qy 136 ttaccgaaataggaataatccttggccacctgtattgcaaggcgcaatgctaatttt 195
      || || || || || || || || || || || || || || || || || || || ||
Db 619 TTTACCTHWKTKTTGCGADAACCTCTCATCTCTTAATTTGKAATCAGCCWTTTTT 560
      || || || || || || || || || || || || || || || || || || || ||
Qy 196 ttcttttccagagctctcaaaaaaaaaaaaaaaaaaccccttactaacaacaggatcc 255
      || || || || || || || || || || || || || || || || || || || ||
Db 559 TTTCTTTTCAACCCCTTTAAAAAATTTTAAAAAATTTTACAGCTTGAATTC 500
      || || || || || || || || || || || || || || || || || || || ||
Qy 256 cggatgtagctcgatgtcccccattaaacggtaatatttcaggcgctccacactaa 315
      || || || || || || || || || || || || || || || || || || || ||
Db 499 AAACAGTTGGAGCTGAGGTGTTTCTTACCTTTGACATGAGACAATTCACCAAGCAA 440
      || || || || || || || || || || || || || || || || || || || ||
Qy 316 t 316
Db 439 T 439

RESULT 9
AL514053/c
LOCUS
DEFINITION
  AL514053 LTI_NFL006_PL2 Homo sapiens cDNA clone CL0BA012ZG07 3
  prime, mRNA sequence.
ACCESSION
  AL514053
VERSION
  AL514053.1 GI:12777547
KEYWORDS
  EST.
SOURCE
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```



```

REFERENCE
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
Location/Qualifiers
1..829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL0BA0122G07"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 616 8371
Email : liliang@lifestech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 213 a 153 c 103 g 247 t 113 others
ORIGIN
Query Match 8.5%; Score 40.8; DB 9; Length 829;
Best Local Similarity 42.4%; Prod. No. 1.3e-02;
Matches 84; Conservative 22; Mismatches 92; Indels 0; Gaps 0;
QY 54 ttttttcaaatagggaatgtttttttcttccctttttataacacctgacccacaga 113
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 202 TTTTITTTGCAAAAGGATTTTGTGATKTTTITKTTTITKTTTITTTTGTGDKA 143
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 114 ctgacgttctaggagcccttaccgaaataagaaataaaccttgccaccttaat 173
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 142 TCCAAATTCAAAGGACATCTCTCTTAAATTCCTTTTATTTTGGATKTTTIT 83
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 174 ttcaaaagcaatgctaatgtttttcttctctccacagctctcaaaaaaataaaaa 233
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 82 GTCAAAATCTTACHTHTHAATTTTWTGTAAGAAAAAATAAAAAAATAAAAAA 23
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 234 accttactaataaacagg 251
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 22 AAAAAAAAAAAAAAAGG 5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 10
CNS01450/c
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence sp6 end of BAC
BACN15M09 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106426
VERSION AL106426.1 GI:5621382
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1201)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk ". This Drosophila melanogaster BAC

```

```

library (Dros BAC) was made by Alain Billard at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
source
Location/Qualifiers
1..1201
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15M09"
/note="End : SP4"
BASE COUNT 302 a 149 c 89 g 474 t 231 others
ORIGIN
Query Match 8.5%; Score 47.6; DP 12; Length 1201;
Best Local Similarity 49.0%; Prod. No. 2.7e-02;
Matches 51; Conservative 19; Mismatches 44; Indels 0; Gaps 0;
QY 143 aaataaqaataaatacttccacaccttattcttgaagaagaataatatttttttt 202
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 867 ATAAATGTAATAAAAAAAMWACCCGTCYCYTITTTAAWIAAAWAAAAAAWTTTTTT 808
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 203 ctcaaaactctctcaaaaadaadaadaaaacacttactatataaa 246
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 807 ATWTAATTTDHAAGAACAAAAAAMAAAVVAAVAAWAA 744
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 11
B1540842
LOCUS
DEFINITION 454279 MARCH 160V Bos taurus cDNA 5', mRNA sequence.
ACCESSION B1540842
VERSION B1540842.1 GI:15361954
KEYWORDS EST.
SOURCE Bos taurus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Smith,I.P.L., Grosse,W.M., Freking,B.A., Peters,S.A.J., Stone,E.L.,
Casas,E., Wray,J.E., White,J., Che,J., Janjekrad,S.C., Bennett
G.L., Heaton,M.P., Loeferlein,W., Kott,R.G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,J., Karanyelova,S., Liard,F., Quakenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4): 626-630 (2001)
Contact: Smith IPI.
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68901-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smiths@mail.mars.usda.gov
Single pass sequencing. Bases called and all trimmed with phred 18
and -mismatch 12 options.
PCR primers
FORWARD: AGGAAACAGTATGACATA;
BACKWARD: GTCTCCACTGACCAATG
Plate: 118 row: M column: 16
Seq primer: ATTATGTGACACTAG.
FEATURES
source
Location/Qualifiers
1..467
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARCH"
/tissue_type="tracheal"
/lab_host="DH10b"

```


[illegible]

Search completed: July 5, 2006
Job time: 679 sec

